

AD \_\_\_\_\_

Award Number: DAMD17-02-1-0073

TITLE: Translational Regulation of PTEN/MMAC1 Expression  
in Prostate Cancer

PRINCIPAL INVESTIGATOR: Jian-Ting Zhang, Ph.D.

CONTRACTING ORGANIZATION: Indiana University  
Indianapolis, IN 46202-5167

REPORT DATE: May 2005

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20051101 063

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> 01/05/2005		<b>2. REPORT TYPE</b> Final		<b>3. DATES COVERED (From - To)</b> 1 May 02 - 30 Apr 05	
<b>4. TITLE AND SUBTITLE</b> Translational Regulation of PTEN/MMAC1 Expression in Prostate Cancer				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b> DAMD17-02-1-0073	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Jian-Ting Zhang, Ph.D.  E-Mail: jianzhan@iupui.edu				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  Indiana University Indianapolis, IN 46202-5167				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for Public Release; Distribution Unlimited.					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b>  In this project, we proposed to use a dicistronic expression system to determine whether the long 5'-UTR sequence of PTEN contains internal ribosome entry site (IRES) which can mediate cap-independent translation. We have accomplished the proposed work as planned. However, we found that the long 5'-UTR sequence of PTEN does not have the predicted IRES activity but rather contains a strong promoter. We have mapped this promoter and it is likely responsible for constitutive production of the PTEN mRNAs with shorter 5'-UTRs in prostate cancer cells which would be compatible for cap-dependent translation initiation. We have also created a promoterless dicistronic vector and validated its use as a more stringent assay for testing cellular IRES activities.					
<b>15. SUBJECT TERMS</b> Prostate cancer, PTEN, apoptosis, IRES, protein synthesis, 5'-UTR, mRNA, translation					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b>
<b>a. REPORT</b> U	<b>b. ABSTRACT</b> U	<b>c. THIS PAGE</b> U			<b>19b. TELEPHONE NUMBER (include area code)</b>
			UU	71	

## Table of Contents

<b>Cover.....</b>	<b>1</b>
<b>SF 298.....</b>	<b>2</b>
<b>Table of Contents .....</b>	<b>3</b>
<b>Introduction.....</b>	<b>4</b>
<b>Body.....</b>	<b>4</b>
<b>Key Research Accomplishments.....</b>	<b>7</b>
<b>Reportable Outcomes.....</b>	<b>7</b>
<b>Conclusions.....</b>	<b>7</b>
<b>References.....</b>	<b>7</b>
<b>Appendices.....</b>	<b>9</b>

## • INTRODUCTION

PTEN/MMAC1/TEP1 (referred as PTEN in the remaining text of this paper) is a tumor suppressor gene that maps to the 10q23.3 and has been shown to be deleted or mutated in many human tumors including glioblastomas, endometrial neoplasms, hematological malignancies, and prostate and breast cancers (Ali et al., 1999; Cantley and Neel, 1999; Dahia, 2000; Di Cristofano and Pandolfi, 2000). In addition, germline mutations in PTEN cause Cowden syndrome (CS), characterized by multiple hamartomas and a high proclivity for developing cancer (Cantley and Neel, 1999). The 5'-UTRs of human and mouse PTEN are highly homologous (the 0.9 kb 5'-UTR near the translation start codon shows ~90% homology) and have many features of a translationally-regulated gene. Firstly, they are significantly longer (>0.95 kb) than the average eukaryotic 5'-UTR. Secondly, they contain multiple short open reading frames (>3 for human and mouse PTEN). Thirdly, they are GC rich with predicted stable complex secondary structures. These features of 5'-UTR are expected to significantly hinder translation initiated by the conventional cap-dependent mechanism. Thus, an alternative mechanism such as internal ribosome entry may be involved in regulating the translation of PTEN. It is possible that in some advanced prostate cancers PTEN expression is omitted due to miss-regulation of IRES-mediated translation initiation. In this project, we plan to test this theory which is novel and has not been anticipated previously.

## BODY

In the original application, we proposed to test the hypothesis that the 5'-UTR of PTEN mRNA contains a critical *cis*-element for IRES-dependent translation and mutation of this element may cause lost expression of PTEN in some prostate tumors. Five tasks were proposed to be accomplished to test this hypothesis.

**Task 1: Cap-dependent translation and IRES activity of PTEN/MMAC1 5'-UTR.** The 5'-UTR of PTEN/MMAC1 will be cloned into pGEM vector for *in vitro* and into a dicistronic vector for *in vivo* analysis.

**Task 2: Boundary of PTEN/MMAC1 IRES and mutagenesis effect.** Systematic deletion and site-directed mutagenesis will be performed using the dicistronic construct to determine the boundary of the IRES of PTEN/MMAC1 5'-UTR and to map the important element.

**Task 3: Sporadic mutations of PTEN/MMAC1 5'-UTR in advanced prostate tumor.** Samples of advanced prostate tumors will be tested first for PTEN/MMAC1 expression at both the RNA and protein level. mRNA will be isolated from the samples that express RNA but not protein of PTEN/MMAC1. cDNA encoding the 5'-UTR of PTEN/MMAC1 will be cloned by 5'-RACE PCR from these mRNA samples and sequenced to identify potential mutations that affect the translation of PTEN/MMAC1.

**Task 4: IRES activity in different cell lines.** The 5'-UTR of PTEN/MMAC1 in the dicistronic construct in Task 1 will be used to determine the IRES activity in various cell lines.

**Task 5: IRES activity during apoptosis.** The 5'-UTR of PTEN/MMAC1 in the dicistronic construct in Task 1 will be introduced into HeLa cells and the IRES activity will be determined under apoptotic conditions.

During the funding period, we have accomplished testing the 5'-UTR sequence of PTEN as planned in the task 1, task 2, task 4, and task 5 as described below.

Analysis of the 5'-UTR using Zuker's mfold program (version 3.0) (Mathews et al., 1999) showed that this region has potential to form various secondary structures and all of the structures predicted are extremely stable with a typical free energy as high as -432 kcal/mole, which is well beyond the energy shown to inhibit ribosome scanning. In addition, this region is also punctuated with four translation start codon AUGs that encode two short upstream open reading frames (uORFs) with 36 and 45 amino acids, respectively (two of the AUGs are located at the end of the first uORF in the same frame). Both uORFs end before -551 bases upstream of the physiological translation start codon. Generally, uAUGs or uORFs inhibit translation of the main ORF by interfering the scanning process of the 40S ribosome complex. To analyze the effect of the 5'-UTR sequence of mouse PTEN on translation, we utilized an *in vitro* transcription and translation system. Mouse PTEN cDNA with and without the 5'-UTR sequence were cloned into pGEM-4Z under the control of SP6 promoter. Run-off transcripts were generated by SP6 RNA polymerase, purified and quantified, and equal amount of RNAs were used to program cell-free translation in rabbit reticulocyte lysate. We found that no full-length PTEN protein

- was translated from transcripts containing the long 5'-UTR whereas significant amount of PTEN protein was generated in the absence of the 5'-UTR sequence. Clearly, the presence of the long 5'-UTR sequence drastically inhibits the translation of PTEN in rabbit reticulocyte lysate.

To analyze whether the protein-encoding sequence (ORF) influence the translational inhibition effect of the 5'-UTR, we cloned the 5'-UTR sequence of PTEN immediately upstream of an open reading frame encoding firefly luciferase in a pGEM-4Z based reporter vector, pG-LUC, resulting in the plasmid pG-5'LUC. This construct was then used to program in vitro transcription and cell-free translation as described above. Again, the translation programmed by pG-LUC transcript generated significantly increasing amount of firefly luciferase activity proportional to the amount of transcripts used while no luciferase activity was detected from the translation programmed by pG-5'LUC transcripts. Thus, the 5'-UTR sequence of mouse PTEN inhibits translation of transcripts encoding a heterologous protein and is independent of the protein-encoding sequence.

The 5'-UTR sequence of PTEN has also been cloned into the dicistronic expression vector pRF and tested for its IRES activity. The IRES sequence of human rhinovirus (HRV) was engineered in the same way and was used as a positive control. The pRF-based constructs contain an SV40 promoter to direct the transcription of dicistronic RNA encoding *Renilla* luciferase as the first cistron and firefly luciferase as the second cistron. Translation of the first cistron (*Renilla* luciferase) serves as an indicator of cap-dependent translation while translation of the second cistron (firefly luciferase) reflects the IRES activity-associated with the inserted intergenic sequence. This approach is considered as a "gold standard" for characterizing cellular IRES (Sachs, 2000). These dicistronic constructs were transfected into HeLa cells and both *Renilla* and firefly luciferase activities were measured. We found that the 5'-UTR sequence of PTEN stimulated the expression of firefly luciferase (second cistron) by ~70 fold over negative vector control and about 2-3 fold over the positive pR-HRV-F control. Similar results were also obtained using another cell line H1299. These results suggest that the 5'-UTR sequence of PTEN may contain an IRES element.

We have also accomplished the systematic deletion and site-directed mutagenesis to determine the boundary of the IRES of PTEN/MMAC1 5'-UTR as planned in task 2. To define the boundaries in the potential IRES in the 5'-UTR sequence of PTEN for enhancing expression of the second cistron, a series of deletion mutants of the 5'-UTR of PTEN were engineered for dicistronic assay. Deletion from both ends of the 5'-UTR resulted in gradual decrease in stimulating firefly luciferase expression. Since the deletion mutant D828-551 and D220-1 retains the most enhancing activity, the 230-nucleotide central region of the 5'-UTR (-551 to -220) may contain the important elements for stimulating firefly luciferase expression.

However, it has been argued that the use of dicistronic vector for determination of IRES activity suffers an inevitable drawback that the 5'-UTR sequence may contain a cryptic promoter (Kozak, 2001; Schneider and Kozak, 2001). In the past, this argument has been essentially ignored. To address this potential problem, we created a promoterless dicistronic vector by deleting the SV40 promoter of the vector. We argued that if the 5'-UTR sequence contains a promoter rather than an IRES, the second cistron would be expressed regardless with or without the SV40 promoter. To validate this newly created vector system, we first tested the well-known IRES of eIF4G (Gan et al., 1998; Gan and Rhoads, 1996). To our surprise, we found that the known IRES of eIF4G is in fact a promoter. This conclusion was supported by several other evidence such as RNA transfection assay and mapping the promoter sequence and identification of transcription factors involved. This study has been published (*Mol. Cell. Biol.* 22:7372-7384, see appendix for detailed description).

We next moved to study the 5'-UTR of PTEN mRNA. We found that the 5'-UTR of PTEN mRNA continued to stimulate the expression of the second cistron in the promoterless dicistronic vector, suggesting that the 5'-UTR of PTEN mRNA may contain a promoter. The existence of this cryptic promoter drives production of a PTEN transcript with shorter 5'-UTRs which can be translated with much higher efficiency. This promoter was mapped to the region between -550 and -220 bases upstream of the translation start codon. Co-transfection analysis using *Drosophila* SL2 cells showed that Sp1 is one of the major transcription factors that can constitutively activate this promoter. Two endogenous PTEN transcripts with 5'-UTRs of 193 and 109

- bases were found in the prostate cancer cell line DU145. Based on these observations, we concluded that the PTEN expression may be regulated at the transcriptional level and that the 5'-UTR sequence of PTEN contains a promoter that is responsible for constitutive PTEN expression. This work has now been published (*Oncogene* 22:5325-5337, see appendix for detailed description).

Because our study proved that the long 5'-UTR of PTEN mRNA does not contain any putative IRES activity as we previously thought, we decided not to pursue further in the direction as we proposed in our original application to test if there are sporadic mutations in the 5'-UTR of PTEN in prostate tumors which may affect the putative IRES activity (task 3). Instead, we analyzed two more well-known IRES elements to validate our findings on the 5'-UTR of PTEN mRNA using promoterless dicistronic vectors. One is from the 5'-UTR of PDGF-B while the other is from p27<sup>kip</sup>. Both PDGF-B and p27 are known to play important roles in prostate cancer. The 5'-UTRs of both PDGF-B and p27 are long and have been thought to have strong IRES activities. The 5'-UTRs of both PDGF-B and p27 were cloned into the intergenic region of the traditional and the promoterless dicistronic vectors and were tested for their putative IRES activities. Our findings are described below.

We found no evidence that the 5'-UTR sequence of PDGF-B mRNA contains any IRES activity. Instead, we found that the 5'-UTR sequence of PDGF-B functions as a promoter both constitutively and upon induction in a variety of cell lines as with PTEN. The 5'-UTR sequence contains two promoters (termed P1 and P2) when only the 5'-UTR sequence is analyzed. In the presence of the upstream TATA-box containing promoter (P0), P1 and P0 promoter are integrated into one promoter while P2 promoter still functions. The full promoter with combined P0, P1 and P2 produced two transcripts, with the major one that has the full-length 5'-UTR and the minor one that has a short 5'-UTR. The integrated P0/P1 promoter and P2 promoter are likely responsible for producing the endogenous 3.8-kb and 2.8-kb PDGF-B mRNAs that are detected in cultured human renal microvascular endothelial cells (HRMECs), a few tumor cells and rat brain tissues. Furthermore, we detected the 2.8-kb PDGF-B mRNA in erythroleukemia K562 cells upon TPA-induced differentiation. Considering that the 5'-UTR in the 3.8-kb mRNA contains no IRES activity and inhibits cap-dependent translation, we believe that the endogenous 2.8-kb mRNA produced from the 5'-UTR promoter is likely the major template responsible for protein production both constitutively and upon induction. We also found that the transcription from the 5'-UTR P2 promoter might be coordinated by the major upstream P0 promoter upon stimulation. Based on these observations, we propose that the TATA-containing P0 promoter and the 5'-UTR promoter work together to tightly control the expression of PDGF-B. This work has been published (*J. Biol. Chem.* 278:46983-46993, see appendix for detailed description).

We also found no evidence that the 5'-UTR sequence p27 mRNA contains any IRES activity using the promoterless dicistronic vector. In the study of p27, we also created the promoterless monocistronic vector to test the promoter activity of p27 and indeed confirmed our findings. The previously observed IRES activities of p27 are, thus, likely generated from the promoter activities present in the 5'-UTR sequences of p27. Therefore, transcriptional regulation may play an important role in p27 expression which has not been anticipated in the past. This finding, together with our studies on eIF4G, PTEN, PDGF-B all indicate that more stringent studies such as promoterless dicistronic and monocistronic DNA and dicistronic RNA tests are required to safeguard any claims of cellular IRES. The manuscript describing this work is under consideration by *Nucleic Acid Research* (see manuscript in appendix for detailed description).

Based on our findings on PTEN, eIF4G, PDGF-B, and p27, we conclude that the prevailing IRES activities of many cellular mRNAs require more stringent testing. Currently, claims on IRES activities of many cellular mRNAs are rapidly increasing and most of these studies do not have the stringent testing such as the use promoterless dicistronic vector, dicistronic mRNA, and monocistronic DNA vector. Our studies cast serious doubts on these claims of cellular mRNAs and help provide better approaches for future studies in the field of translational control.

## • KEY RESEARCH ACCOMPLISHMENTS

1. Created the promoterless dicistronic vector as a more stringent method for testing cellular IRES activities.
2. Validated the promoterless dicistronic vector system and tested the known cellular IRES of eIF4G.
3. Demonstrated the previously known cellular IRES from eIF4G was misled by its promoter activities.
4. Cloned the 5'-UTR of PTEN mRNA into the pGEM vectors and investigated the effect of the long 5'-UTR of PTEN on translation.
5. Found that the long 5'-UTR of PTEN mRNA inhibits translation of PTEN mRNA in vitro. Removal of the 5'-UTR sequence drastically enhanced the translation efficiency.
6. Cloned the 5'-UTR of PTEN mRNA into the promoterless dicistronic assay system for IRES assay and found that the 5'-UTR of PTEN does not have an IRES as previously thought and it contains cryptic promoters which is responsible for constitutive PTEN expression in prostate cancer cell lines.
7. The promoter in the 5'-UTR of PTEN has been mapped to the region between -550 and -220 bases upstream of the translation start codon.
8. Sp1 transcription factor was found to be the major factor responsible for internal transcription initiation of PTEN.
9. Changed direction and further validated the dicistronic vector system for translational control studies.
10. Found no evidence for the previously known cellular IRES from the 5'-UTRs of PDGF-B and p27 which are important molecules for prostate cancers.

## REPORTABLE OUTCOMES

1. Han, B.; **Zhang, J.T.** Regulation of gene expression by internal ribosome entry site (IRES) vs cryptic promoter: the eIF4G story. *Mol. Cell. Biol.* 22:7372-7384; 2002.
2. Han, B.; Dong, Z.; Liu, Y.; Chen, Q.; Hashimoto, K.; **Zhang, J.T.** Regulation of constitutive expression of mouse PTEN by the 5'-untranslated region. *Oncogene* 22:5325-5337; 2003.
3. Han, B.; Dong, Z.; **Zhang, J.T.** Tight control of platelet-derived growth factor B (PDGF-B/c-sis) expression by interplay between the 5'-untranslated region sequence and the major upstream promoter. *J. Biol. Chem.* 278:46983-46993; 2003.
4. Liu, Z.\*; Dong, Z.\*; Han, B.; Liu, Y.; **Zhang, J.T.** Regulation of expression by promoters vs internal ribosome entry site in the 5'-untranslated sequence of the human cyclin-dependent kinase inhibitor p27<sup>kip1</sup>. *Nucleic Acid Res.* (under revision)

## CONCLUSIONS

In conclusion, the long 5'-UTR sequence of PTEN inhibits translation of PTEN mRNAs and it contains promoters that mediate production of PTEN mRNAs with shorter 5'-UTRs which are likely responsible for constitutive production of PTEN protein. The mRNAs of PTEN derived from these promoters have shorter 5'-UTRs which are compatible with the 5'-cap dependent translation initiation and ribosome scanning mechanism. The long 5'-UTR of PTEN likely does not have an IRES element as we proposed originally.

We have created a more stringent promoterless dicistronic testing system for cellular IRES. Three of the previously claimed cellular IRESs (eIF4G, PDGF-B, and p27) have now been disproved using our more stringent system. Thus, the past studies on many cellular IRES elements may contain major flaws and a new test using the promoterless dicistronic vector needs to be used to safeguard any future claims of cellular IRES activities. We believe that this series of our studies have very high impact on the future studies of translational control in general and on that of cellular IRES specifically.

## REFERENCES:

Ali, I. U., Schriml, L. M., and Dean, M. (1999). Mutational spectra of PTEN/MMAC1 gene: a tumor suppressor with lipid phosphatase activity. *J Natl Cancer Inst* 91, 1922-1932.

- Cantley, L. C., and Neel, B. G. (1999). New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. *Proc Natl Acad Sci U S A* 96, 4240-4245.
- Dahia, P. L. (2000). PTEN, a unique tumor suppressor gene. *Endocr Relat Cancer* 7, 115-129.
- Di Cristofano, A., and Pandolfi, P. P. (2000). The multiple roles of PTEN in tumor suppression. *Cell* 100, 387-390.
- Gan, W., LaCelle, M., and Rhoads, R. E. (1998). Functional characterization of the internal ribosome entry site of eIF4G mRNA. *J Biol Chem* 273, 5006-5012.
- Gan, W., and Rhoads, R. E. (1996). Internal initiation of translation directed by the 5'-untranslated region of the mRNA for eIF4G, a factor involved in the picornavirus-induced switch from cap-dependent to internal initiation. *J Biol Chem* 271, 623-626.
- Kozak, M. (2001). New ways of initiating translation in eukaryotes? *Mol Cell Biol* 21, 1899-1907.
- Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288, 911-940.
- Sachs, A. B. (2000). Cell cycle-dependent translation initiation: IRES elements prevail. *Cell* 101, 243-245.
- Schneider, R., and Kozak, M. (2001). New Ways of Initiating Translation in Eukaryotes? *Mol Cell Biol* 21, 8238-8246.



## Regulation of Gene Expression by Internal Ribosome Entry Sites or Cryptic Promoters: the eIF4G Story

Baoguang Han and Jian-Ting Zhang\*

*Department of Pharmacology and Toxicology, Walther Oncology Center/Walther Cancer Institute and I.U. Cancer Center, Indiana University School of Medicine, Indianapolis, Indiana 46202*

Received 20 February 2002/Returned for modification 29 April 2002/Accepted 19 July 2002

As an alternative to the scanning mechanism of initiation, the direct-internal-initiation mechanism postulates that the translational machinery assembles at the AUG start codon without traversing the entire 5' untranslated region (5'-UTR) of the mRNA. Although the existence of internal ribosome entry sites (IRESs) in viral mRNAs is considered to be well established, the existence of IRESs in cellular mRNAs has recently been challenged, in part because when testing is carried out using a conventional dicistronic vector, Northern blot analyses might not be sensitive enough to detect low levels of monocistronic transcripts derived via a cryptic promoter or splice site. To address this concern, we created a new promoterless dicistronic vector to test the putative IRES derived from the 5'-UTR of an mRNA that encodes the translation initiation factor eIF4G. Our analysis of this 5'-UTR sequence unexpectedly revealed a strong promoter. The activity of the internal promoter relies on the integrity of a polypyrimidine tract (PPT) sequence that had been identified as an essential component of the IRES. The PPT sequence overlaps with a binding site for transcription factor C/EBP $\beta$ . Two other transcription factors, Sp1 and Ets, were also found to bind to and mediate expression from the promoter in the 5'-UTR of eIF4G mRNA. The biological significance of the internal promoter in the eIF4G mRNA might lie in the production of an N-terminally truncated form of the protein. Consistent with the idea that the cryptic promoter we identified underlies the previously reported IRES activity, we found no evidence of IRES function when a dicistronic mRNA containing the eIF4G sequence was translated *in vitro* or *in vivo*. Using the promoterless dicistronic vector, we also found promoter activities in the long 5'-UTRs of human Sno and mouse Bad mRNAs although monocistronic transcripts were not detectable on Northern blot analyses. The promoterless dicistronic vector might therefore prove useful in future studies to examine more rigorously the claim that there is IRES activity in cellular mRNAs.

Translational control plays an important role in regulating gene expression in eukaryotic cells (9, 24, 30, 52, 54, 69, 73). Most of the translational regulation occurs at the level of initiation, which is usually the rate-limiting step in protein synthesis (15, 29, 44, 50, 68). Initiation of translation in eukaryotes normally depends on the m<sup>7</sup>GpppN 5'-cap structure of mRNA, which recruits the 43S ribosome preinitiation complex via interaction with the cap binding protein eIF4E (69). The translational machinery then migrates downstream until it meets the first AUG codon in the optimal context (41, 43). This scanning model implies that the migration of the 43S preinitiation complex is inhibited by any long and highly structured 5'-untranslated regions (5'-UTRs). About 90% of 5'-UTRs of vertebrate mRNAs have a length in the range between 100 and 300 bases, which is compatible with the ribosome-scanning model. However, the remaining mRNAs have longer atypical 5'-UTRs which may also contain one or more AUG codons. These mRNAs may not be efficiently translated using the cap-dependent initiation and ribosome-scanning mechanism (73). Furthermore, cap-dependent translation initiation is compromised under many cellular conditions including mitosis (5), apoptosis (10, 49), cellular stresses

such as hypoxia and heat shock (55, 60, 64), and viral infection (45, 53, 56).

Under the aforementioned cellular conditions, many proteins still need to be synthesized possibly via the internal ribosome entry site (IRES)-mediated initiation mechanism (36). In this mechanism, the ribosome preinitiation complex is recruited directly to the IRES element independent of the 5'-cap structure and begins to scan within the vicinity of the initiator AUG codon by effectively avoiding a large part of the 5'-UTR sequence. The IRES-mediated initiation was first demonstrated a decade ago in picornaviruses (3, 19, 27, 71). Picornavirus mRNAs are naturally uncapped at their 5' ends. Their 5'-UTRs usually have significant secondary structure spanning approximately over 500 bases and are punctuated by multiple AUG codons. Such a structure is predicted to inhibit the ribosome recruitment and linear ribosome scanning. However, these 5'-UTRs can effectively confer a cap-independent translation by directly recruiting ribosomes.

Recently, it has been suggested that some cellular mRNAs may also use the IRES-mediated initiation mechanism (27, 61, 71). Indeed, over 40 mRNAs have been reported to use IRESs to direct translation initiation, and the list is growing rapidly. An IRES database (<http://www.rangueil.inserm.fr/IRESdatabase>) has been created to reflect their potential importance and to categorize the cellular IRES elements. However, the existence of such cellular IRESs has been questioned recently due to the inevitable drawbacks of the conventional dicistronic test and the fact that dicistronic mRNA assays do not work for the

\* Corresponding author. Mailing address: Department of Pharmacology and Toxicology, IUCC, Indiana University School of Medicine, 1044 W. Walnut St., R4-166, Indianapolis, IN 46202. Phone: (317) 278-4503. Fax: (317) 274-8046. E-mail: jianzhan@iupui.edu.

known cellular IRESs (42). This has generated debate over the existence of cellular IRESs and how studies should be conducted on future cellular IRES candidates (63). In most of the previous cellular IRES studies, an artificial dicistronic construct was used in which a candidate IRES element is inserted between the two cistrons. The upstream cistron acts as an indicator of cap-dependent translation, and the downstream cistron indicates the IRES activity. This approach is considered a "gold standard" for characterizing cellular IRESs (61). However, it is acknowledged by both sides of the debate that one of the major questions is whether cryptic promoter or potential splice sites exist in the "cellular IRES" sequence of the dicistronic construct that may generate monocistronic mRNA of the second cistron. Although RNA analyses such as Northern blotting were conducted to rule out the existence of monocistronic mRNA in most previous studies. Kozak (42, 63) argued that the RNA analysis approaches used in these studies were not sensitive enough to rule out the existence of monocistronic mRNA present at 5% or less of the dicistronic mRNA level. Thus, some of the previously reported cellular IRESs might be falsely identified, and it is clear that relevant rigorous experimental methods are needed to clarify these disputes.

In this study, we created a new promoterless dicistronic vector to rule out the existence of cryptic promoters in any potential cellular IRES elements. This analysis also relies on enzyme assays in the same way as the conventional dicistronic test for IRES activity and therefore is more compatible with IRES study than any existing RNA analysis approaches. Using this vector, we first tested the 5'-UTR of eukaryotic translation initiation factor 4G1 (eIF4G), the largest subunit of eIF4F. eIF4G is important for translational control by acting as a molecular scaffold and coordinating the activities of eIF3, eIF4A, eIF4E, and poly(a)-binding protein (PABP) (22, 28, 39). The translation of eIF4G is cap independent under conditions of viral infection (37) and involves an IRES-dependent mechanism (20, 21, 37). The 5'-UTR of eIF4G has been functionally characterized and shown to have a polypyrimidine tract (PPT) region that is essential for IRES activity (20, 21).

Using the novel promoterless dicistronic vector, we demonstrated that a strong promoter activity exists in the 5'-UTR sequence of eIF4G that can explain the enhanced expression of the second cistron observed in the conventional dicistronic tests. Further mapping of the promoter sequence in the 5'-UTR of eIF4G showed that the PPT sequence was absolutely required for the promoter activity. Using gel shift and super-shift analysis, we showed that transcription factor C/EBP $\beta$  binds to this PPT sequence. In addition, the 5'-UTR of eIF4G could not direct IRES-dependent translation in rabbit reticulocyte lysate (RRL) or in HeLa cells when dicistronic RNA transcripts were analyzed. Thus, the cryptic promoter in the 5'-UTR of eIF4G underlies the enhanced expression of the second cistron observed in the conventional dicistronic test. Using the same promoterless dicistronic vector, we also observed promoter activities in the 5'-UTR sequences of human Sno and mouse Bad, although no monocistronic transcripts were detected by northern blot. Hence, cryptic promoter activities in the 5'-UTR sequences of cellular mRNAs are more prevalent than anticipated, and they can be misinterpreted to be IRES sequences in the conventional dicistronic analysis due to the low intrinsic sensitivity of Northern blot analysis for

ruling out the existence of monocistronic transcripts. We propose that the promoterless dicistronic vector may be used to establish claims of cellular IRES in future studies.

## MATERIALS AND METHODS

**Materials.** Restriction enzymes, m<sup>7</sup>GpppG cap analogue, *Pfu* polymerase, and SL2 cell line were purchased from New England Biolabs, Amersham/Pharmacia Biotech, Stratagene, and the American Type Culture Collection, respectively. T7 and Sp6 RNA polymerases, RNasin, RNase-free DNase, rabbit reticulocyte lysate, Luciferase assay 'Stop & Glo' kit, pGL3-Promoter plasmid, and pSP64 poly(A) plasmid were from Promega. RNeasy Mini Kit and Oligotex mRNA Mini Kit were from Qiagen. Rediprime II Random Prime Labeling System and [<sup>32</sup>P]-dCTP were from Amersham Biosciences. The Sephadex G-25 Quick Spin Columns (TE) for radiolabeled DNA purification were from Roche. MAGNA nylon transfer membrane was from Osmonics Inc. Wild-type and mutant C/EBP $\beta$  oligonucleotides used in the electrophoretic mobility shift assay (EMSA) and antibodies to nuclear proteins Sp1, Sp3, and C/EBP $\beta$  were obtained from Santa Cruz Biotechnology. Schneider's *Drosophila* culture medium, Lipofectamine Plus, and Lipofectin transfection reagent were purchased from Invitrogen. Oligonucleotides were synthesized by Sigma-Genosys. I.M.A.G.E. expressed sequence tag EST clones were obtained from either the American Type Culture Collection or Research Genetics.

**Construction of plasmids.** To engineer dicistronic constructs containing the IRES element of human rhinovirus (HRV) or the 5'-UTR of eIF4G, the dicistronic vector pRF (70) was used. Plasmid pGL2/CAT/4G/LUC containing the 5'-UTR of eIF4G (20) was used as template for PCR to amplify the 5'-UTR of eIF4G by using the following two primers containing *SpeI* and *NcoI* sites: 5'-CAAAGTCTAGTCTAGATGGGGGTCT-3' and 5'-CAACCATGGTGATATCCTTCTCC-3'. The PCR product was cloned into the *SpeI* and *NcoI* sites of pRF vector to obtain plasmid pR-eIF4G-F (see Fig. 1A). The IRES sequence of HRV in pGL3-R-HRV plasmid (70) was released by digestion with *SpeI* and *NcoI* and cloned into pRF, resulting in pR-HRV-F.

To engineer promoterless dicistronic constructs that allow analysis of promoter activity of the DNA insert in the intergenic region, the simian virus 40 (SV40) promoter sequence including the chimeric intron between *SmaI* and *EcoRV* was removed from pRF, pR-eIF4G-F, and pR-HRV-F (see Fig. 1A), resulting in pRF(-P), pR-eIF4G-F(-P), and pR-HRV-F(-P), respectively. For all the promoter analysis studies, the promoterless dicistronic constructs were used. In some studies, the SV40 enhancer was deleted by cloning the *NheI*-*HpaI* fragment from the promoterless dicistronic plasmid pR-eIF4G-F(-P) into the pGL3-promoter vector (Promega), which contains an SV40 poly(A) signal but does not contain the SV40 enhancer. The resulting construct is named pR-eIF4G-F(-PE). Systematic deletions of the 5'-UTR sequence of eIF4G were generated by PCR. The eIF4G linker-scanning constructs were made by the PCR-based overlap extension technique, which is similar to the PCR-based site-directed mutagenesis method described previously (25). A 10-base linker with the sequence ACTCTAGACT was used to replace wild-type sequences.

To generate poly(A)-tailed in vitro transcripts for RNA transfection study, constructs containing poly(A) were engineered using the vector pSP64 PolyA (Promega), which has a 30-bp (dA-dT) sequence. The *EcoRV*-*XbaI* fragment of pRF vector that contains the *Renilla* luciferase gene was first cloned into the pSP64 PolyA vector at the *XbaI* and blunted *HindIII* sites to generate plasmid pSP-R<sub>A30</sub>. The *XbaI* fragment of pR-HRV-F that contains the IRES of HRV and the firefly luciferase gene were then isolated and cloned into pSP-R<sub>A30</sub> at the *XbaI* site to generate pSP-R-HRV-F<sub>A30</sub>. The pSP-RF<sub>A30</sub> plasmid was obtained by removing the IRES sequence of HRV from pSP-R-HRV-F<sub>A30</sub> by digestion with *SpeI* and *NcoI*. To engineer pSP-R-eIF4G-F<sub>A30</sub>, the 5'-UTR sequence (from -368 to -44) of eIF4G that has been reported to contain full IRES activity (20) was amplified by PCR using the following primers with *SpeI* and *NcoI* sites: 5'-CAAAGTCTAGTCTAGATGGGGGTCT-3' and 5'-ACACCATG GATTCGGATCTGGGA-3'. The PCR products were used to replace the IRES sequence of HRV in pSP-R-HRV-F<sub>A30</sub>.

To clone the 5'-UTRs of mouse Bad, human Sno, and human inhibitor apoptosis 1 (hiAP-1, referred to hereafter as HIAP) in both pRF and pRF(-P) vectors, I.M.A.G.E. EST clones were used as templates. The 405-bp 5'-UTR sequence of mouse Bad (accession no. NM\_007522) was amplified using mouse I.M.A.G.E. clone 948554 (accession no. AA544696) with the primers 5'-GACTGGTCACACCTATCCTGGCA-3' and 5'-GCCATGGTTGGATCCTGGAAGCCTG-3'. The 709-bp 5'-UTR sequence of human Sno (accession no. U70730) was assembled using I.M.A.G.E. clones 1713010 (accession no. A1129968) and 211285 (accession no. H69011), and finally amplified by the

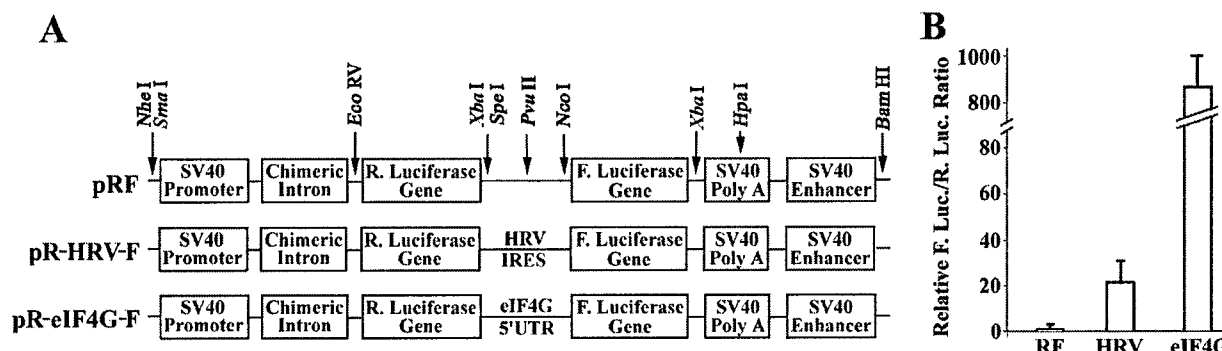


FIG. 1. Stimulation of the second-cistron expression by the 5'-UTR sequence of eIF4G. (A) Schematic diagram of dicistronic constructs without insert (pRF), or with the IRES of HRV (pR-HRV-F) and the 5'-UTR of eIF4G (pR-eIF4G-F) in the intergenic region. The locations of several relevant restriction enzyme sites are shown by arrows. (B) Relative luciferase activity generated by the dicistronic constructs. HeLa cells were transfected with pRF, pR-HRV-F, and pR-eIF4G-F constructs. At 24 h following transfection, the cells were harvested, the *Renilla* and firefly luciferase (R. Luc. and F. Luc.) activities were measured, and the relative ratios were calculated and normalized to that of the vector-transfected cells (RF). The data are from six independent experiments.

following two primers: 5'-AACTAGTGGTTTCAAATTGGCCCT-3' and 5'-G GTTTTCCATGGTACACTCT-3'. The 1.3-kb 5'-UTR sequence of human HIAP (accession no. AF070674) was assembled using human I.M.A.G.E. clones 2349482 (accession no. A1827488) and 428231 (accession no. AA002125) by a combination of PCR and restriction digestion. All of the 5'-UTR cDNA fragments were finally cloned between the *Spe*I and *Nco*I sites of the pRF and pRF(-P) vectors, resulting in pR-Bad-F, pR-Sno-F, pR-HIAP-F, pR-Bad-F(-P) and pR-Sno-F(-P). All the above constructs were confirmed by DNA sequencing.

**Cell culture and DNA and RNA transfection.** HeLa and H1299 cells were maintained at 37°C under 5% CO<sub>2</sub> in Dulbecco modified Eagle medium and RPMI 1640 medium, respectively, supplemented with 10% fetal bovine serum. Schneider's *Drosophila* cell line 2 (SL2) was maintained at room temperature under atmospheric CO<sub>2</sub> in Schneider's *Drosophila* medium supplemented with 10% fetal bovine serum.

**DNA transfection in both HeLa and H1299 cells** was performed with Lipofectamine Plus reagents as specified by the manufacturer. In a 24-well plate, approximately 10<sup>5</sup> cells/well were plated and were transfected with 0.4 µg of DNA the next day. Cells were harvested for the luciferase assay 24 h following transfection. SL2 cells were transfected using the calcium phosphate method as previously described (25). Briefly, 4 µg of pR-eIF4G-F(-P) or pR-eIF4G-F(-PE) was cotransfected with 2 µg of *Drosophila* expression plasmids for Sp1, Ets-1 (38), or combination of Sp1 plus Ets-1. Equal amounts of DNA were obtained in each transfection by adding empty vector PacO. Cells were incubated with calcium phosphate-precipitated DNA for 48 h and then harvested for the luciferase assay.

**RNA transfection** was performed using the cationic liposome-mediated method as described by Dwark et al (16). Briefly, approximately 2 × 10<sup>5</sup> cells/well were seeded onto six-well plates on the day before transfection. Opti-MEM I medium (1 ml) in a 12-by 75-mm polystyrene snap-cap tube was mixed with 12.5 µg of Lipofectin reagent and 5 µg of capped mRNA. The liposome-RNA-medium mixture was immediately added to cells. At 8 h following transfection, the cells were harvested and processed for luciferase analysis.

**Preparation of cytoplasmic extract (S100).** Cytoplasmic extract (S100) was prepared as previously described (31). Briefly, HeLa cells were collected using rubber policeman, washed with phosphate-buffered saline and then resuspended at a density of 5 × 10<sup>7</sup> cells/ml in homogenization buffer H100 (10 mM Tris-HCl [pH 7.4], 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.5 mM dithiothreitol, 0.1 mM phenylmethylsulfonyl fluoride, 10 µg of leupeptin per ml). The cells were homogenized with a Dounce homogenizer for 20 strokes on ice. Cell nuclei were pelleted by centrifugation at 2,000 × g for 10 min. The supernatant was then adjusted to 150 mM KCl and centrifuged at 100,000 × g at 4°C for 90 min. The supernatant was recovered as S100 extract. The extract was flash frozen with 5% glycerol and stored at -70°C.

**In vitro transcription and translation.** In vitro transcription and translation were performed as previously described (75). DNA templates with a poly(A) tail were linearized with *Eco*RI, while DNA templates without a poly(A) tail were linearized with *Bam*HI. Capped transcripts were synthesized in the presence of 1 mM m<sup>7</sup>GpppG and purified using the Qiagen RNeasy mini kit. A 50-ng portion

of capped RNA transcripts was used to program cell-free translation in rabbit reticulocyte lysate (RRL) in a final volume of 10 µl containing 6.5 µl of RRL. The translation mixture contained 2 µl of either H100 buffer or cytoplasmic extract S100 (8.0 µg/µl).

**Nuclear extract preparation, EMSA, and UV cross-linking.** HeLa nuclear extract preparation, EMSA, and supershift analysis were done as previously described (25). The two strands of oligonucleotides used as EMSA probes were annealed prior to labeling. The sequences of the sense strand of the oligonucleotide probes are 5'-TAGCTTTCTTCCCGAGATCC-3' (eIF4G C/EBPβ [-68 to -48]), 5'-GAGGTGGGCTCTTCCTGCTTCC-3' (eIF4G Ets-1 [-101 to -80]), and 5'-GCTGGGGGGTGGGGAGTTGG-3' (eIF4G Sp1 [-151 to -132]). The oligonucleotides used for competition analysis are wild-type MRG1 Sp1 (5'-TTAAGCTTCGCTCCGCCCTTCC-3'), mutant MRG1 Sp1 (5'-TTAAGCTTCGCTTCGCCCTTCC-3') (25), wild-type C/EBPβ (5'-TGCAGATTGC GCAATGTGCA-3'), and mutant C/EBPβ (5'-TGCAGAGACTAGTCTCTGC A-3'). For UV cross-linking, the reaction mixture containing probe and nuclear extract was incubated at room temperature for 20 min and then irradiated with UV at 254 nm at a distance of 5 cm for 30 min by using a UV Stratilinker 1800 (Stratagene). The mixture was then separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (10% polyacrylamide).

**Northern blot analysis.** Subconfluent H1299 cells in 10-cm plates were transfected with constructs (4 µg/plate) using Lipofectamine Plus. The total RNAs were extracted using an RNeasy mini kit 48 h following transfection. Residual plasmid DNA in the total RNA was digested with RNase-free DNase. The poly(A) RNAs were then isolated from 250 µg of total RNAs using the Oligotex mRNA mini kit. One-fifth of the mRNAs were separated in 1% agarose gels in the presence of formaldehyde and morpholinepropanesulfonic acid (MOPS) buffer and blotted onto MAGNA nylon membranes. The blots were hybridized with a <sup>32</sup>P-labeled firefly luciferase DNA probe (1,656 bp), which was isolated from pRF by cleaving with *Nco*I and *Xba*I and labeled using Rediprime II random-prime labeling system.

## RESULTS

**The 5'-UTR of eIF4G directs expression of the second cistron in a dicistronic construct.** To further analyze the IRES of eIF4G, we reengineered the dicistronic plasmids by cloning the cDNA encoding the 5'-UTR of eIF4G into a widely used dicistronic vector, pRF (70) (see also Fig. 1A). This vector contains the SV40 promoter to direct the transcription of dicistronic RNA encoding *Renilla* luciferase as the first cistron and firefly luciferase as the second cistron. Following the SV40 poly(A) signal, there is an SV40 enhancer that enhances the promoter activity. The IRES sequence of HRV was engineered in the same way as eIF4G and was used as a control (Fig. 1A). These plasmids were transfected into HeLa cells, and both

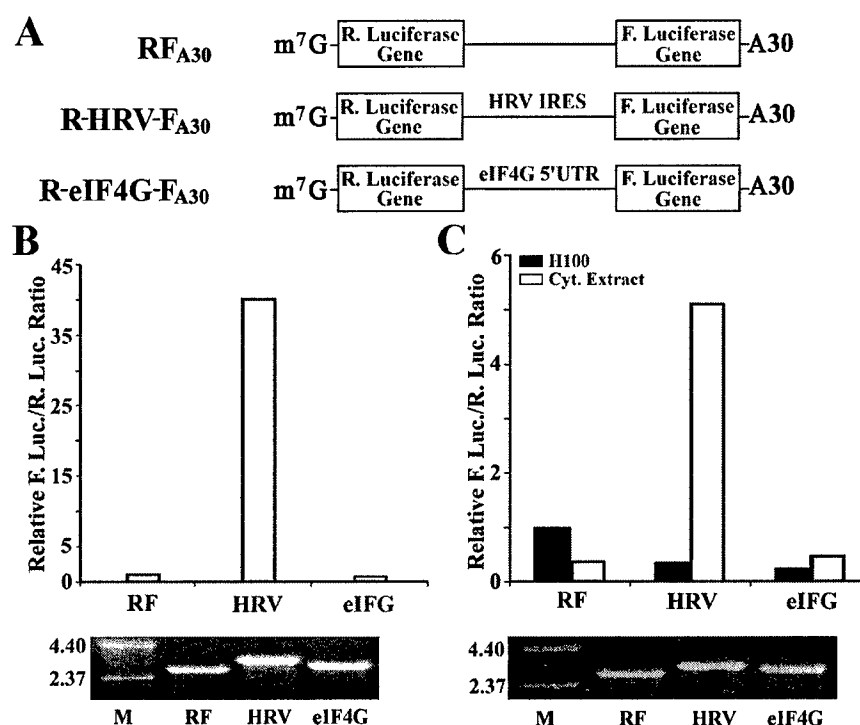


FIG. 2. Translation of dicistronic mRNA in HeLa cells and in RRL. (A) Schematic diagram of the dicistronic mRNA used for translation in HeLa cells. In vitro transcripts with 5'-cap ( $m^7G$ ) and 3'-poly(A) tail ( $A_{30}$ ) were synthesized using T7 RNA polymerase from linearized vector alone ( $RF_{A30}$ ), constructs containing the IRES of HRV ( $R-HRV-F_{A30}$ ), and the 5'-UTR of eIF4G ( $R-eIF4G-F_{A30}$ ). R., *Renilla*; F., firefly. (B) Relative luciferase activity from dicistronic mRNAs in HeLa cells. HeLa cells were transfected with the dicistronic mRNAs, and 8 h following transfection, *Renilla* and firefly luciferase (R. Luc. and F. Luc.) activities were measured and the relative ratios were calculated and normalized to that of the vector-transfected cells (RF). (C) Relative luciferase activity from dicistronic RNA in RRL. Capped dicistronic transcripts were synthesized using T7 RNA polymerase from linearized pRF, pR-HRV-F, and pR-eIF4G-F (Fig. 1). The in vitro transcripts were then used to program translation in RRL in the presence of buffer alone (H100) or the HeLa cytoplasmic extract (S100). Following the cessation of translation, *Renilla* and firefly luciferase activities were measured, the ratio of firefly to *Renilla* activity was determined, and the relative ratios were calculated and normalized to the vector control in the presence of buffer (H100). The data are representative of three independent experiments. The gels shown at the bottom are in vitro transcripts (500 ng each) separated on a 1% agarose gel that were used for in vitro and in vivo studies, respectively.

*Renilla* and firefly luciferase activities were measured. As shown in Fig. 1B, the 5'-UTR of eIF4G displays an unusually high activity in directing the expression of firefly luciferase. The activity is ~850-fold higher than that of the control vector pRF and ~40-fold higher than that of the HRV IRES. Similar results were observed after transfecting other cell lines such as H1299 cells (data not shown). This observation is consistent with previous studies which suggested that the stimulation of the second cistron expression was due to IRES activities in the 5'-UTR of eIF4G (20, 21).

**Translation of dicistronic RNA transcripts in HeLa cells and RRL.** The above results in Fig. 1 suggest that (i) the 5'-UTR of eIF4G contains IRES activity which enhances the translation of firefly luciferase from the dicistronic mRNA by internal initiation, as suggested previously (20, 21); (ii) the 5'-UTR of eIF4G may contain a cryptic promoter which directs transcription of the firefly luciferase gene; and/or (iii) the 5'-UTR of eIF4G contains a cryptic splicing acceptor site which creates a splicing variant with only the second cistron of the firefly luciferase gene. To distinguish between these possibilities, we first generated dicistronic RNAs in vitro from the dicistronic constructs and used them to program translation both in HeLa cells and in RRL. RNA transfection is one of the

major methods for characterizing translation efficiency and identifying eukaryotic regulatory factors influencing IRES activity. This method bypasses the complex issue of transcriptional regulation and requires only the cytoplasmic delivery of the transcripts. For purposes of RNA transfection, pSP- $RF_{A30}$ , pSP-R-eIF4G- $F_{A30}$ , and pSP-R-HRV- $F_{A30}$  were engineered and used for producing dicistronic transcripts containing the  $m^7GpppG$  cap and polyadenylated tail in vitro (Fig. 2A). Transcripts were introduced into HeLa cells by Lipofectin encapsulation (16). At 8 h following transfection, cell lysates were prepared for luciferase activity measurement. As expected, the firefly luciferase of the vector RNA was very poorly translated and its activity (arbitrary units) represented only about 0.16% of that of *Renilla* luciferase (data not shown). It increased to about 6.7% with the dicistronic RNAs containing HRV IRES. Therefore, HRV IRES significantly stimulated the translation of firefly luciferase, about 40-fold higher than that of the vector control (Fig. 2B). However, no stimulation of firefly luciferase expression was observed with the 5'-UTR of eIF4G. We next used the dicistronic RNA to program translation in RRL. As shown in Fig. 2C, the IRES of HRV did not show any enhancement of firefly luciferase expression in RRL (solid bar), indicating that RRL may lack factors necessary for

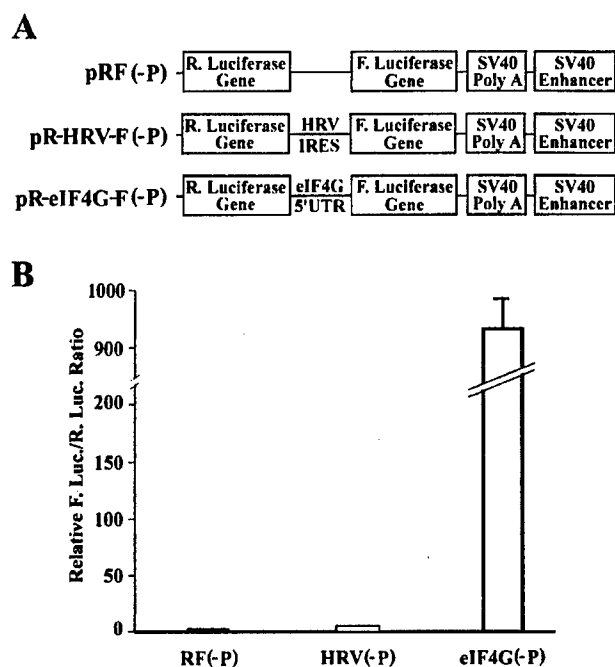


FIG. 3. Cryptic promoter activity of the 5'-UTR of eIF4G. (A) Schematic diagram of promoterless dicistronic construct of pRF(-P), pR-HRV-F(-P) and pR-eIF4G-F(-P). The sequences of the SV40 promoter and chimeric intron were removed from the parental dicistronic constructs shown in Fig. 1A. (B) Relative luciferase activity generated from the promoterless constructs. The promoterless constructs were transfected into HeLa cells, and 24 h following the transfection, cells were harvested for determination of *Renilla* and firefly luciferase (R. Luc. and F. Luc.) activity. The relative ratios between firefly and *Renilla* luciferase activities were calculated and normalized to that of the vector-transfected cells [pRF(-P)]. The data were from six independent experiments performed in duplicate.

IRES-dependent translation initiation. The addition of HeLa S100 cytoplasmic extract to RRL stimulated ~40-fold expression of firefly luciferase under the control of IRES of HRV (Fig. 2C). However, no stimulation of firefly luciferase expression was observed by the 5'-UTR of eIF4G in either the absence or presence of HeLa S100 extract (Fig. 2C), confirming the results of RNA transfection studies. Thus, the 5'-UTR of eIF4G may not contain an IRES element to direct cap-independent translation from a dicistronic RNA transcript.

**Cryptic promoter activity in the 5'-UTR of eIF4G.** To analyze whether the 5'-UTR of eIF4G contains any promoter activity, we simply removed the unique SV40 promoter together with the intron sequence from the dicistronic constructs. These promoterless dicistronic constructs (Fig. 3A) were then transfected into HeLa cells for determination of both *Renilla* and firefly luciferase activities (Fig. 3B). As expected, both the *Renilla* and firefly luciferase activities were minimal but could be detected for the pRF(-P) vector control. Only a twofold increase in firefly luciferase activity was observed with the pR-HRV-F(-P) construct. This small increase is in dramatic contrast to the 20-fold increase when the pR-HRV-F construct was used, as shown in Fig. 1B. Thus, the enhanced expression of firefly luciferase from pR-HRV-F constructs (Fig. 1) was not due to production of monocistronic

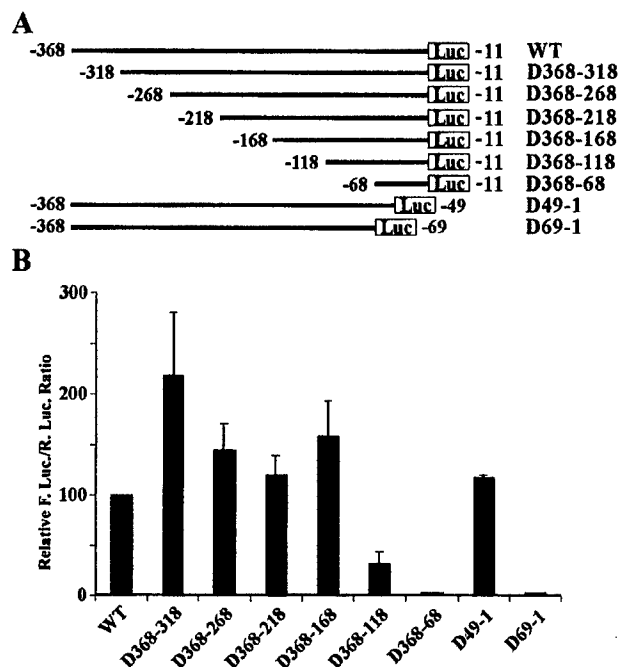


FIG. 4. Effects of deletions on the promoter activity of the 5'-UTR of eIF4G. (A) Schematic diagram of the deletions in the 5'-UTR of eIF4G. The positions of the 5' and 3' ends of each deletion are indicated on the left and right, respectively. These mutant 5'-UTRs were engineered into the promoterless dicistronic vector at the intergenic region. (B) Relative luciferase activity generated from the wild-type and mutant 5'-UTR sequences of eIF4G. HeLa cells were transfected with the constructs shown in panel A, and 24 h following transfection, *Renilla* and firefly luciferase (R. Luc. and F. Luc.) activities were measured and the ratio of firefly to *Renilla* luciferase was calculated and normalized to that of the wild type (WT) control.

transcript by a cryptic promoter in the HRV sequence. Surprisingly, the pR-eIF4G-F(-P) construct generated more than 900-fold-higher firefly luciferase activity than that of the vector control. This increase was similar to that generated by the pR-eIF4G-F construct shown in Fig. 1B, suggesting that the significant enhancement in the firefly luciferase expression by the 5'-UTR sequence of eIF4G is probably due to the presence of a strong promoter in this region.

**Mapping of the eIF4G promoter in the 5'-UTR.** The above studies suggest that the 5'-UTR sequence of eIF4G contain a cryptic promoter which may have been thought to be an IRES element in previous studies. To further validate our study, we mapped the cryptic promoter by deleting DNA sequences systematically from both the 5' and 3' ends of the 5'-UTR of eIF4G in the promoterless dicistronic vector (Fig. 4A). As shown in Fig 4B, the promoter activity increased significantly when 50 bases was deleted from the 5' end of the 5'-UTR (D368-318), indicating that there is a repressor element within this region. The promoter activity did not significantly change after further deletion up to 200 bases from the 5' end (D368-168). However, deletion of an additional 50 bases from the 5' end resulted in a dramatic decrease in promoter activity (D368-118), suggesting that the 168 bases upstream of the translation start site contains the full promoter activity of the 5'-UTR of eIF4G. Furthermore, deletion of the region from

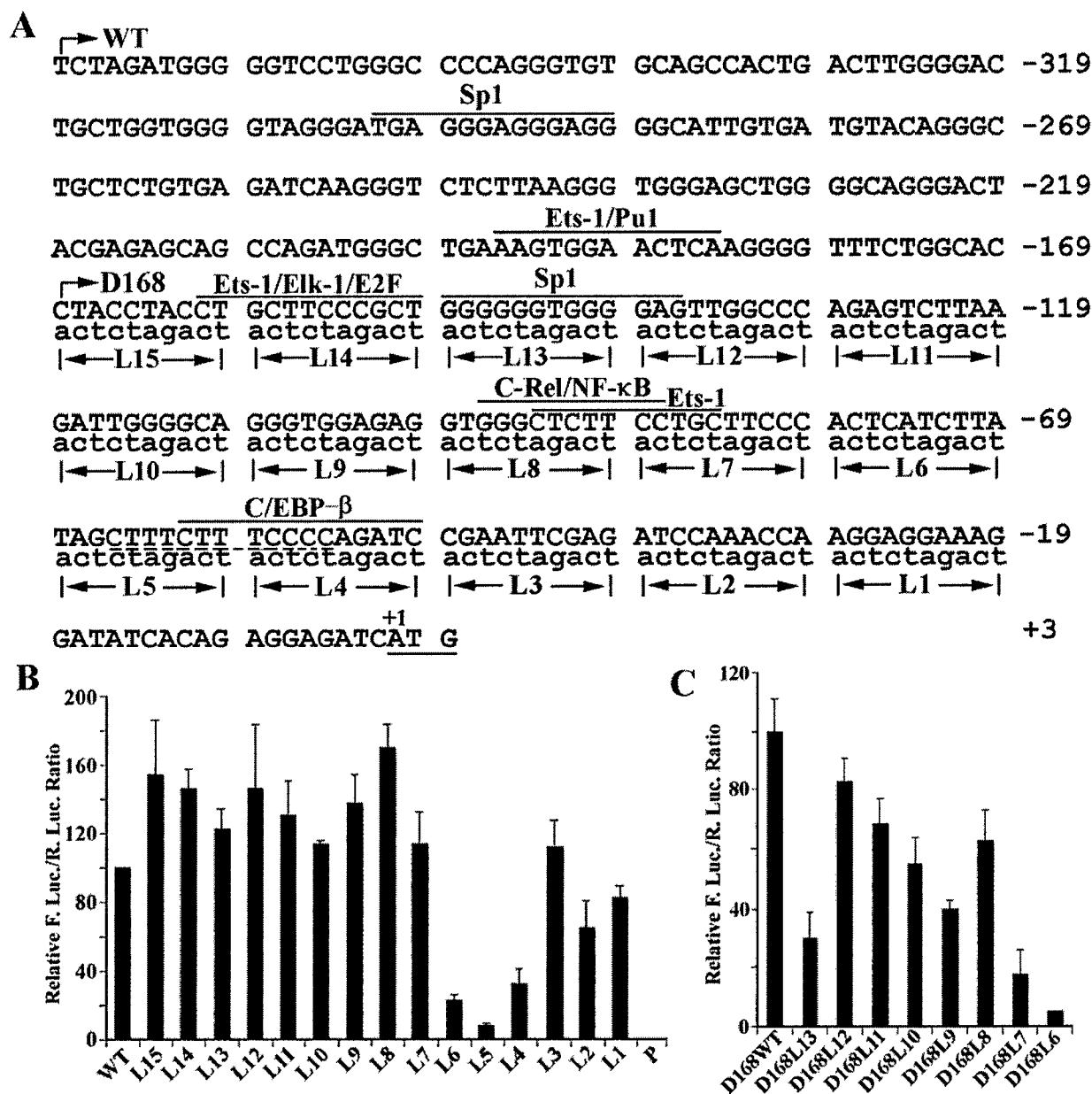


FIG. 5. Mapping of sequences important for the promoter activity in the 5'-UTR of eIF4G. (A) Schematic demonstration of the linker-scanning mutations. Linker-scanning mutations were made onto the full-length or truncated D168 constructs. The putative transcription factor binding sites predicted by MatInspector software are shown above the sequence. The dashed line indicates the polypyrimidine tract region important for IRES activity. WT, wild type. (B and C) Relative luciferase activity of wild-type and mutant full-length (B) or D368-168 (C) constructs in HeLa cells. At 24 h following transfection of the constructs into HeLa cells, the cells were harvested for determination of *Renilla* and firefly luciferase activity. The ratio of firefly to *Renilla* luciferase activity was calculated and normalized to that of the wild-type sequences. Data are from three independent experiments.

–49 to –1 with respect to the 3' end did not significantly affect promoter activity, whereas deletion of the region from –69 to –1 resulted in complete loss of the promoter activity. Thus, the region from –69 to –49 upstream of the translation start site probably contains the critical elements for the promoter in the 5'-UTR sequence of eIF4G. Interestingly, this region corresponds to the PPT region that was shown previously to be critical for the IRES activity of eIF4G (20, 21) (see Fig. 5A for

the sequence).

**Linker-scanning analysis of the cryptic promoter in the 5'-UTR of eIF4G.** To identify potential transcription factor binding sites required for the cryptic promoter activity in the 5'-UTR of eIF4G, we engineered a series of linker-scanning constructs near the translation start site. Linker-scanning constructs were engineered into the full-length 5'-UTR sequence by site-directed mutagenesis in which 10 successive nucleotides

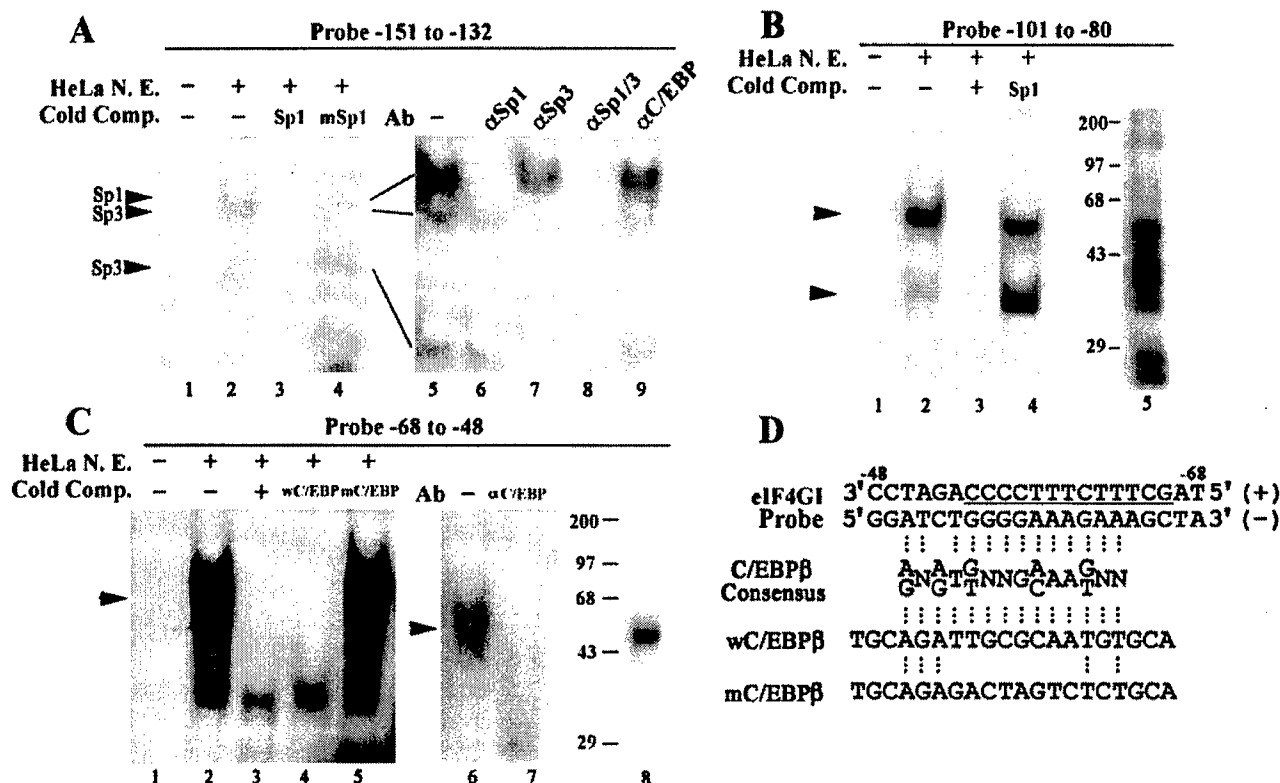


FIG. 6. Identification of nuclear proteins binding to the regions from -151 to -132, -101 to -80, and -68 to -48 in the 5'-UTR of eIF4G. (A) EMSA and supershift analysis of a probe with the sequence from -151 to -132. HeLa nuclear extract (N. E.) (10  $\mu$ g) was incubated with the probe alone (lane 2) or in the presence of unlabeled Sp1 (lane 3), mutant Sp1 (lane 4) oligonucleotide as competitors or in the absence (lane 5) or presence of anti-Sp1 (lane 6), anti-Sp3 (lane 7), both anti-Sp1 and anti-Sp3 (lane 8), or anti-C/EBP $\beta$  (lane 9) antibodies (Ab) for the supershift assay. The DNA-protein complexes were separated by PAGE. The DNA-protein complexes generated are indicated by arrows. The gel of lanes 5 to 9 was run longer than the gel of lanes 1 to 4 to better separate the protein complexes. (B) EMSA and UV cross-linking of a probe with the sequence from -101 to -80. HeLa nuclear extract (10  $\mu$ g) was incubated with the probe alone (lane 2) or in the presence of the unlabeled probe (lane 3) or Sp1 oligonucleotide (lane 4) as competitors. Cross-linking of DNA-protein complexes was performed by incubating 10  $\mu$ g of nuclear extract with the labeled probe followed by UV irradiation and separation by SDS-PAGE (lane 5). (C) EMSA, supershift, and UV cross-linking of a probe with the sequence from -68 to -48. HeLa nuclear extract (10  $\mu$ g) was incubated with the probe alone (lane 2), or in the presence of the unlabeled probe (lane 3), wild-type C/EBP $\beta$  (lane 4) or mutant C/EBP $\beta$  (lane 5) oligonucleotides as competitors, or in the absence (lane 6) or presence (lane 7) of anti-C/EBP $\beta$  antibody for the supershift assay. The DNA-protein complexes were separated by PAGE. UV cross-linking was performed in the way same as described for panel B, with the probe from -68 to -48. (D) Sequence comparison of the consensus C/EBP $\beta$  binding site with the probe from -68 to -48. Also shown are the wild-type (w) and mutant (m) C/EBP $\beta$  oligonucleotides used for the EMSA in panel C. Conserved nucleotides are aligned by three dots. N = A, C, G, or T. The PPT of eIF4G is underlined. Lane 1 in all experiments (panels A to C) is a control consisting of the probe alone incubated without nuclear extract.

were replaced with the sequence ACTCTAGACT (Fig. 5A). This sequence was chosen because it contains no sequence homologous to any known transcription factor binding sites. These constructs were then transfected into HeLa cells, and the promoter activity was measured using cell lysate following transfection. As shown in Fig. 5B, replacement with L4, L5, or L6 (corresponding to -79 to -49 upstream of the translation start site) resulted in significant decreases in promoter activity. This observation was consistent with the results of the deletion-mapping study shown in Fig. 4B (construct D69-1). However, the results in Fig. 4B (compare the results for constructs D368-168, D368-118, and D368-68) also suggested that this region (-79 to -49) itself did not contain the full promoter activity. To identify other *cis* elements, we next engineered the same linker-scanning constructs into the D368-168 truncated constructs since it contains minimal sequence for the full promoter

activity. As shown in Fig. 5C, two additional regions, -149 to -139 (L13) and -89 to -69 (L7 and L6), were shown to be important for the proximal promoter activity.

**Identification of transcription factors for the cryptic promoter in the 5'-UTR of eIF4G.** The above results suggest that there are several critical regions in the cryptic promoter of the 5'-UTR of eIF4G. To determine the biochemical composition of protein complexes binding to these regions, nuclear extracts from HeLa cells were analyzed by EMSA using labeled probes corresponding to regions from -151 to -132, -101 to -80, and -68 to -48. These sequences were chosen because they are overlapped with the critical regions identified by promoter-mapping analysis shown in Fig. 4 and 5. In addition, sequence analysis (57) showed that they contain Sp1/Sp3, Ets-1, and C/EBP $\beta$  binding sites, respectively (Fig. 5A). As shown in Fig. 6A, the probe from -151 to -132 formed three protein com-

plexes in EMSA (lanes 2 and 5). Addition of cold Sp1 oligonucleotides prevented the formation of the three complexes (lane 3). In contrast, addition of cold mutant Sp1 oligonucleotides had no effect on the complex formation (lane 4). Furthermore, addition of Sp1 antibody to the reaction mixture inhibited the formation of the uppermost complex (lane 6) while addition of Sp3 antibody inhibited the formation of the two lower complexes (lane 7). Addition of both antibodies completely inhibited the formation of all three complexes (lane 8), while addition of C/EBP $\beta$  antibody had no effect (lane 9). These results clearly indicate that the region from -151 to -132 contains a typical Sp1/Sp3 binding site.

The region from -101 to -80 was predicted to contain an Ets-1 binding sequence (Fig. 5A). EMSA analysis showed that two specific protein complexes were formed (Fig. 6B, lane 2). Addition of cold probe inhibited the formation of these complexes (lane 3), while addition of nonrelevant cold Sp1 probe had no effect on complex formation (lane 4). As shown by UV cross-linking (lane 5), at least six major proteins with different molecular weights bound to this sequence, indicating that this sequence is recognized by a large number of different proteins. The identities of these proteins remain to be determined.

The region from -68 to -48 was predicted to contain a C/EBP $\beta$  binding sequence. As shown in Fig. 6D, C/EBP $\beta$  binding sequence involves 12 bases with a core sequence of (A/G)N(A/G)T(T/G)NNG(A/C)AA(T/G). The homologous sequence of C/EBP- $\beta$  in the 5'-UTR of eIF4G is located in the region from -61 to -50 that overlaps the PPT. EMSA analysis showed that a single specific protein complex was formed with this probe (Fig. 6C, lane 2). This complex is specific, since addition of cold probe prevented the formation of this complex (lane 3). The addition of cold probe corresponding to the consensus C/EBP $\beta$  binding sequence also competed for the formation of the complex (lane 4). In contrast, a cold mutant C/EBP $\beta$  probe had no effect on complex formation (lane 5). Hence, the protein involved in binding to the sequences of -68 to -48 is probably C/EBP $\beta$ . This conclusion was further confirmed using C/EBP $\beta$  antibody, which inhibited the formation of the complex (lane 7). UV cross-linking analysis also showed a single protein with the same molecular weight as C/EBP $\beta$  (lane 8). Taken together, these results demonstrated that the region from -68 to -48 contains a C/EBP $\beta$  binding sequence.

**Transcription factors Sp1 and Ets-1 synergistically transactivate the activity of eIF4G promoter in SL2 cells.** To determine directly whether the above transcription factors could functionally modulate the promoter activity in the 5'-UTR sequence of eIF4G, *Drosophila* SL2 cells, which are deficient in Sp1-, Sp3-, and Ets-related proteins (12, 13), were used. The reason for using insect instead of mammalian cells is that Sp1-, Sp3-, and Ets-related factors are expressed in virtually all mammalian cells, which could affect the interpretation of this experiment. We introduced the pR-eIF4G-F(-P) construct along with *Drosophila* expression plasmid pPacSp1 or pPacUEts-1 into *Drosophila* SL2 cells. As shown in Fig. 7A, both pPacSp1 and pPacUEts-1 stimulated the promoter activity in the 5'-UTR sequence of eIF4G. To test the possible functional interplay between the Sp1- and Ets-related proteins, we performed cotransfection experiments using combinations of plasmids expressing Ets-1 and Sp1. As shown in Fig. 7A, pPacSp1 and pPacUEts-1 synergistically stimulated the promoter activ-

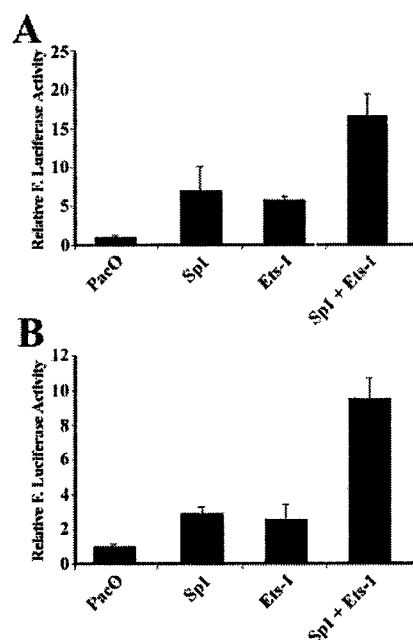


FIG. 7. Sp1 and Ets-1 proteins activate the promoter in the 5'-UTR sequence of eIF4G in *Drosophila* SL2 cells. pR-eIF4G-F(-P) (A) and pR-eIF4G-F(-PE) (B) were transfected into SL2 cells together with *Drosophila* expression vector (PacO), or plasmids expressing transcription factors Sp1 and Ets-1. At 48 h following transfection, the cells were harvested for determination of firefly (F.) luciferase activity, which was normalized to that of vector-transfected cells. The data are from three independent experiments.

ity in the 5'-UTR sequence of eIF4G. Since our promoterless construct pR-eIF4G-F(-P) contains an SV40 enhancer downstream of the poly(A) signal sequence that may complicate the interpretation of the above results (Fig. 3A), we removed the SV40 enhancer region from pR-eIF4G-F(-P) by cloning its *NheI*-*HpaI* fragment into the pGL3-promoter vector, which contains intact SV40 poly(A) signal but without SV40 enhancer. The same pattern of results was observed (Fig. 7B). Thus, the *trans* activation of promoter by Sp1 and Ets-1 occurs through the *cis* elements present in the 5'-UTR sequence of eIF4G.

**Analysis of 5'-UTRs of Sno, Bad, and HIAP by conventional and promoterless dicistronic test.** To determine how robust the promoterless dicistronic test is in differentiating cryptic promoters from IRES activities, we analyzed three additional cellular 5'-UTRs by using both conventional and promoterless dicistronic tests (Fig. 8A). The *sno* gene encodes a component of the histone deacetylase complex and can act as a tumor suppressor in mice (67), although paradoxically it also promotes oncogenic transformation of chicken embryo fibroblasts and differentiation of quail embryo fibroblasts (6). The human *sno* gene has a 5'-UTR of 700 bases that is rich in AT. Bad is a member of the Bcl-2 family of proapoptotic proteins that is thought to exert a death-promoting effect by blocking the pro-survival function of Bcl-X<sub>L</sub> through its heterodimerization with Bcl-X<sub>L</sub> (7). The mouse Bad mRNA has a 5'-UTR of 470 bases which contains a small upstream open reading frame encoding a peptide of 60 amino acids. Human HIAP is a member of the



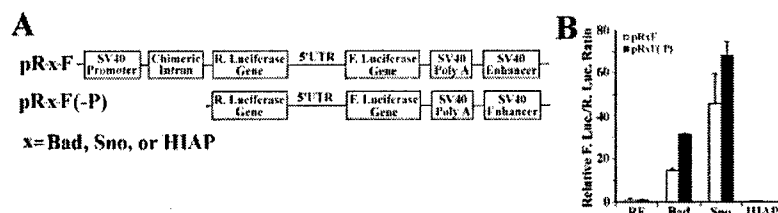


FIG. 8. Conventional and promoterless dicistronic test of the 5'-UTRs of Sno, Bad, and HIAP. (A) Schematic diagram of conventional and promoterless dicistronic constructs of the 5'-UTRs of Sno, Bad and HIAP. (B) Conventional dicistronic and promoterless dicistronic constructs were transfected into HeLa cells. At 24 h following transfection, cells were harvested for enzyme assay and the ratio of *Renilla* (R.) to firefly (F.) luciferase activity was determined and normalized to that of the respective vector controls. The data are from three independent experiments.

IAP protein family that is evolutionally conserved and can block apoptosis when expressed in cells derived from multiple tissues (14). HIAP mRNA has a 5'-UTR of ~3 kb, which would severely inhibit cap-dependent translation initiation (33).

As shown in Fig. 8B, the 5'-UTRs of human Sno and mouse Bad both significantly increased the expression of the second cistron in the conventional dicistronic test (open bars), suggesting that these 5'-UTRs have either IRES and/or cryptic promoter activity. In contrast, no increase in the second-cistron expression was observed with the 5'-UTR of human HIAP, indicating that it does not contain either IRES or cryptic promoter activity. Interestingly, the 5'-UTRs of Sno and Bad stimulated even higher expression of the second cistron in the promoterless dicistronic vector (solid bars). This observation strongly argues that the enhanced expression of the second cistron observed in the conventional dicistronic test is due to the promoter rather than IRES activity in the 5'-UTRs of Sno and Bad.

**Northern blot analysis.** To determine whether the transcript derived from the cryptic promoter of eIF4G, Sno, and Bad can be detected by Northern blotting using the firefly luciferase gene as a probe, poly(A) RNAs were isolated for Northern blot analysis 48 h following transfection of dicistronic constructs pRF, pR-HRV-F, pR-eIF4G-F, pR-Sno-F, pR-Bad-F, and pR-eIF4G-F(-P). As a control for monocistronic transcript, the pRF(-R) construct was engineered by removing DNA sequences encoding *Renilla* luciferase (the 0.95-kb *EcoRV-PvuII* fragment [Fig 1A]). This plasmid uses the SV40 promoter to direct the synthesis of the firefly luciferase transcript without the *Renilla* luciferase sequence. As shown in Fig 9A, the dicistronic transcript from control pRF (lane 1) and the monocistronic transcript from control pRF(-R) (lane 4) were detected as expected. The transcript produced from pR-eIF4G-F(-P) has a size similar to that from pRF(-R) (compare lanes 4 and 7) as expected, suggesting that it is a monocistronic mRNA. The same monocistronic transcript was also observed with pR-eIF4G-F (lane 3). These observations confirm that the 5'-UTR of eIF4G has a strong promoter activity for expressing the second cistron of firefly luciferase. Surprisingly, no dicistronic transcript was found with pR-eIF4G-F (lane 3).

To determine whether the lack of intact dicistronic mRNA from pR-eIF4G-F was due to the low sensitivity of Northern blot analysis, lysates were prepared for the *Renilla* luciferase activity assay from the same cells that were used for the North-

ern blot analysis in Fig. 9A. As shown in Fig. 9B, pR-eIF4G-F displayed about 12% of the *Renilla* luciferase activity of pRF or pR-HRV-F (compare column 3 with columns 1 and 2), suggesting that the level of the dicistronic transcript containing *Renilla* luciferase gene is relatively low. This low level of *Renilla* luciferase activity was consistently observed for pR-eIF4G-F, and the low level of the dicistronic transcript is probably due to alternative splicing events as proposed previously (see Discussion). This event would splice out *Renilla* luciferase sequences from the dicistronic transcripts and generate a monocistronic transcript that encodes only firefly luciferase. The minor transcript, with a size slightly larger than the monocistronic transcript generated from pR-eIF4G-F (Fig. 9A, lane 3), may represent the alternatively spliced products.

Intact dicistronic mRNAs were produced from pR-HRV-F

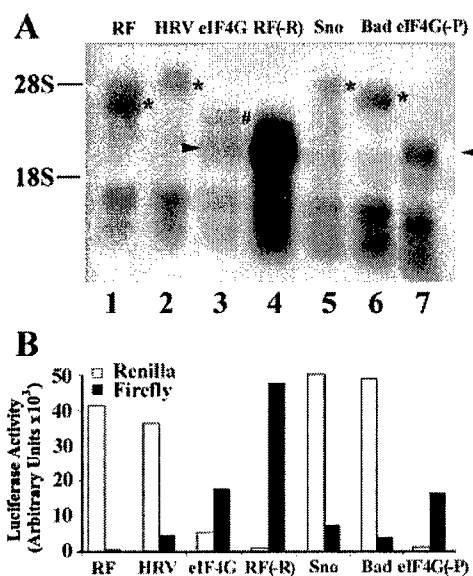


FIG. 9. Northern blot analysis of eIF4G, Sno, and Bad. (A) Poly(A) RNAs were isolated following transfection with pRF (lane 1), pR-HRV-F (lane 2), pR-eIF4G-F (lane 3), pRF(-F) (lane 4), pR-Sno-F (lane 5), pR-Bad-F (lane 6), and pR-eIF4G-F(-P) (lane 7) and used for Northern blot analysis as described in Materials and Methods. The asterisks indicate the dicistronic transcript, the arrowhead indicate the monocistronic transcript and the pound sign indicates the minor transcript. (B) Small fraction of the cells transfected with various constructs for Northern analysis in panel A were used to prepare cell lysates for determination of both firefly and luciferase activities.

(lane 2), pR-Sno-F (lane 5), and pR-Bad-F (lane 6) with the expected sizes. No monocistronic transcripts were detected with these constructs. These observations are consistent with the conclusion that the HRV sequence is an IRES element that does not contain a cryptic promoter. However, the lack of detectable monocistronic transcript from Sno and Bad would have suggested that they, too, do not have cryptic promoters, had we not performed the promoterless dicistronic test (Fig. 8). To determine whether the firefly luciferase was expressed in the cells used for Northern blot analysis, we measured the firefly luciferase activity. As shown in Fig. 9B, the firefly luciferase activity of Sno and Bad is, respectively, ~16- and ~9-fold higher than that generated from pRF and similar to that of pR-HRV-F. Based on these observations, we conclude that Northern blot analysis is not sensitive enough to detect monocistronic transcripts generated from moderate cryptic promoters and thus the promoterless dicistronic assay is a better approach to confirm the existence of IRES in 5'-UTR sequences. It is noteworthy that two unknown bands of high mobility were also detected in all samples on Northern blotting. The similar products have also been observed previously (11). However, because they are smaller than the monocistronic mRNA, they are not expected to be translated into full-length firefly luciferase proteins and do not affect the interpretation of the data.

## DISCUSSION

The recent rapid increase in the number of cellular IRES elements discovered has raised great interest in the field of gene regulation research (27, 61, 71). However, because the conventional dicistronic test inherits inevitable drawbacks and the dicistronic RNA assays do not work for cellular IRES elements, a debate still exists about the validity of cellular IRESs, although viral IRESs are well established (36, 42, 63). In this study, we created and tested a promoterless dicistronic vector which can be used easily to safeguard the claims of cellular IRES in future studies.

One of the more than 40 cellular mRNAs that have been shown to have IRES activities to date is eIF4G (see <http://www.rangueil.inserm.fr/IRESdatabase>). As a major subunit of eukaryotic translation initiation machinery, eIF4G plays a major role in cell growth regulation and cell transformation (2, 18, 26). However, regulation of eIF4G expression is poorly understood. Two different isoforms of cDNA for eIF4G have been isolated (35, 37, 74). The first sequence has a 5'-UTR of 368 bases which contains an active IRES (20, 21). The IRES activity of this 5'-UTR has been well characterized and mapped to a polypyrimidine tract upstream of the AUG start codon (20). The second cDNA sequence (extended eIF4G) encodes a protein that has 156 more amino acids at the amino terminus than the first one and contains a new 5'-UTR, which also contains IRES activity, although the activity is much lower than that of the first 5'-UTR (35, 37). Interestingly, both the truncated and the extended versions of eIF4G function similarly in promoting protein synthesis and cell transformation (18, 26). Currently, it is not known which protein isoforms are expressed in cells. The 5'-UTR of the truncated eIF4G (74) was thought to be an intron of the extended eIF4G gene (23, 37). Analysis of the human genome sequence shows that the 5'-UTR is

located in intron 4 of the extended eIF4G gene (unpublished observation). In this study, we showed that the 5'-UTR of the truncated eIF4G clearly contains promoter sequence which may be used to generate transcripts for expression of truncated eIF4G. It is therefore tempting to speculate that at least two alternative promoters may exist before exon 5 to generate transcripts with various length of the 5'-UTR responsible for production of the truncated eIF4G protein. Indeed, sequence analysis of the eIF4G genome showed that there may be another promoter in the GC-rich region about 1.6 kb upstream of the 5'-UTR of truncated eIF4G (unpublished observation).

The promoter activity associated with the 5'-UTR of eIF4G in this study was probably thought previously to be an IRES activity. First, expression of firefly luciferase of the second cistron driven by the 5'-UTR of eIF4G in a promoterless dicistronic test is at the same level as that observed in a conventional dicistronic test. Second, the cryptic promoter sequence has been mapped and the transcription factor binding sites in the DNA encoding the 5'-UTR have been delineated. Third, some of the transcription factors which are responsible for promoter activation of the 5'-UTR of eIF4G have also been identified using EMSA and complementation assays in insect cells. Fourth, the Northern blot analysis demonstrated that monocistronic transcripts encoding firefly luciferase were produced from both the conventional and the promoterless dicistronic plasmid. Finally, as another alternative strategy to bypass the complex issue of transcription or splicing possibly presented by 5'-UTRs in the conventional dicistronic DNA test, we also performed RNA transfection and *in vitro* translation using dicistronic RNA transcripts. However, the 5'-UTR of eIF4G failed to show any IRES activity. In contrast, the IRES of HRV in dicistronic RNAs can direct the translation of the second cistron both in HeLa cells and in RRL supplemented with HeLa extract. Although similar observations have been made with some other cellular IRESs in previous studies (8, 70), this observation significantly undermines the claim of cellular IRES activity despite the possibility that nuclear experience of the transcripts other than transcription or splicing may contribute to the prerequisites for internal initiation mediated by the cellular IRESs (70).

Interestingly, we showed that the promoter activity in the 5'-UTR sequence of eIF4G relies on the existence of the region from -68 to -48, which corresponds to the PPT that was previously described to be essential for IRES activity of eIF4G. Either deletion or point mutation of this region resulted in dramatic loss of the promoter activity. This region binds to transcription factor C/EBP $\beta$ , a member of the CCAAT/enhancer binding protein (C/EBP) family that is expressed in proliferating cells (46, 51). Although essential, the C/EBP $\beta$  binding element itself displays minimal promoter activity. We also showed by deletion and linker-scanning analysis that the regions from -168 to -68, -151 to -132, and -101 to -80 are also important elements for this promoter activity. While we demonstrated that the region from -151 to -132 binds the Sp1/Sp3 transcription factor, more work is needed to identify the proteins that bind to the region from -101 to -80, which was shown to be highly homologous to the Ets-1 consensus binding sequence. This region binds to at least six proteins in HeLa nuclear extracts, which remain to be identified. Since different Ets proteins exhibit low selectivity in binding-

site preference (66, 72) and since Ets-1 proteins do not exist in *Drosophila* SL2 cells, we analyzed the possible effect of Ets-1 on the promoter activity in the 5'-UTR sequence of eIF4G in SL2 cells. Our results clearly demonstrated that Ets-1 could *trans*-activate the promoter activity in the 5'-UTR sequence of eIF4G. We also demonstrated that Sp1 and Ets-1 work synergistically on this promoter and stimulate the promoter activity by about 10 to 20-fold. The apparent disparity between this number and the nearly 1,000-fold expression noted in Fig. 2 suggests that other transcription factors, C/EBP $\beta$  in particular, may also play a major role in the activation of the 5'-UTR promoter. Based on the results of our promoter analysis, we conclude that the transcription factors C/EBP $\beta$ , Sp1/Sp3, and Ets-1 functionally interact to *trans*-activate the cryptic promoter in the 5'-UTR sequence of eIF4G and therefore may be responsible for the transcriptional regulation of the truncated version of eIF4G.

Interestingly, pR-eIF4G-F generated an additional minor monocistronic transcript (Fig. 9A) that is slightly larger than the normal monocistronic transcript (Fig. 9A), generated by the cryptic promoter in pR-eIF4G-F(-P). This transcript is unlikely to be derived from transcription by the promoter in the 5'-UTR sequence of eIF4G because it is not produced by pR-eIF4G-F(-P). It is probably generated by other mechanisms related to the 5'-UTR sequence of eIF4G because such a transcript was not observed with the vector control and other constructs containing the 5'-UTR of HRV, Bad, and Sno. One such possible mechanism is alternative splicing. The PPT region in the 5'-UTR of eIF4G is followed by a perfect splicing acceptor sequence (TTTCTTTCCCCAGA) which has been suggested to be used for splicing to generate the extended form of eIF4G (23). Use of this acceptor sequence and the upstream donor sequence in the chimeric intron (see Fig. 1) would result in the production of a transcript without the *Renilla* luciferase gene sequence. Furthermore, although the dicistronic transcript was not detected by Northern blot analysis, pR-eIF4G-F produced about 10% *Renilla* luciferase activity compared to other constructs that produced the dicistronic transcripts (Fig. 9), suggesting that about 10% of the dicistronic transcripts exist in the cells transfected with pR-eIF4G-F. The other 90% of the dicistronic transcripts may have been processed to smaller transcripts by alternative splicing. Based on the intensity of the bands on Northern blots (Fig. 9A), the minor monocistronic transcript produced by alternative splicing (lane 3) is about one-third as common as that produced by transcription (lane 3) using the promoter in the 5'-UTR sequence of eIF4G. Therefore, alternative splicing may represent another mechanism responsible for the apparently high firefly luciferase activity observed with the 5'-UTR sequence of eIF4G in the conventional dicistronic test. In light of these findings, it is imperative to rule out the alternative splicing as well as the cryptic promoter when examining a candidate IRES of cellular mRNAs, possibly by testing with an intronless dicistronic construct.

Heterogeneity of transcription initiation has been documented previously for some proto-oncogenes such as *c-myc* (47, 48). Transcription of *c-myc* involves P0, P1, P2, and P3 promoters. Differential promoter usage has been observed in a variety of cell types, and atypical usage of the promoter has generally been associated with abnormal or deregulated con-

trol of cell growth. The transcription starting point of P0, P1, and P2 *c-myc* mRNAs are located 1,172, 524, and 363 bases upstream from the translation start codon CUG, respectively, while that of P3 resides in an intron. P1 and P2 promoter sequences are therefore located in the 5'-UTR of P0 transcripts. Similarly, the 1,038 bases of the 5'-UTR sequence of vascular endothelial growth factor were also demonstrated to contain an alternative transcription initiation site (1).

While it is recognized that the long 5'-UTRs present at the majority of the proto-oncogenes inhibit cap-dependent translation (73) and that IRES-mediated translation initiation may be used for these mRNAs, it has not been well appreciated that these long 5'-UTR DNA sequences may contain promoters for transcription of a less abundant mRNA species with significantly shorter 5'-UTRs. These alternative transcripts with shorter 5'-UTRs are compatible with cap-dependent translation initiation and thus do not require an IRES-mediated translation mechanism. For example, the major transcript of platelet-derived growth factor B chain (PDGF B/*c-sis*) has a 5'-UTR of 1,023 bases that is transcribed by an upstream TATA-containing promoter (17, 40, 58). However, this transcript cannot be translated due to the large size of the 5'-UTR (34, 58, 59). Recently, it has been suggested that IRES-mediated translation initiation may play an important role in the translational regulation of PDGF B/*c-sis* mRNA during cell differentiation (4), and the 630-base sequence within the central portion of the 5'-UTR is important for the IRES activity (65). Interestingly, earlier studies of the PDGF B/*c-sis* gene showed that an additional transcript with a 5'-UTR of only 15 bases was produced in cultured cells on phorbol myristate acetate or transforming growth factor  $\beta$ 1 stimulation (17) and in developing rat brain (62). The production of this shorter transcript was postulated to derive from internal transcription initiation, as suggested by DNase I hypersensitivity analysis of the first exon sequences (17). Sequence analysis indeed revealed multiple Sp1 binding sites in the 5'-UTR sequence of PDGF B/*c-sis* (reference 63 and unpublished observations). It is therefore possible that the alternative transcription for a shorter and less abundant 5'-UTR may be important in regulating PDGF B/*c-sis* expression during cell differentiation, which would undermine the reported IRES activity of the long 5'-UTR of PDGF B/*c-sis*. This possibility certainly merits attention in future studies of the translational regulation of any long cellular 5'-UTRs.

The promoterless dicistronic vector created in this study is clearly a very sensitive approach to detect any potential cryptic promoters in long cellular 5'-UTRs. Promoterless vector has been used in one earlier study for investigation of possible IRES activity in the 5'-UTR from XIAP (X-linked inhibitor of apoptosis protein) (32). The advantage of our vector is that it was derived from parental dicistronic vectors, and therefore data from two sets of vectors are readily comparable. As we have also shown with the 5'-UTRs of Bad, Sno, and HIAP, this vector should be applicable to any cellular 5'-UTRs and can be used as a good control for analysis of cellular IRES elements. Despite the differences in nucleotide composition, the 5'-UTRs of Bad, Sno, and HIAP are all long and punctuated with multiple AUGs that are incompatible with the cap-dependent ribosome-scanning mechanism of translation initiation; therefore, they are potential candidates as IRES elements. How-

ever, using the promoterless dicistronic vector, we found that the 5'-UTRs of Bad and Sno contain promoter activity at a moderate level with no IRES activity whereas the 5'-UTR of HIAP does not contain either promoter or IRES activities. This observation, together with the findings of eIF4G in this study, argues that cryptic promoter activities prevail in the 5'-UTR of cellular mRNAs, which may generate alternative transcripts with shorter 5'-UTRs compatible with cap-dependent translation initiation. Furthermore, our failure to detect the less abundant monocistronic transcripts generated from the promoter in the 5'-UTR sequence of Bad and Sno by Northern blot analysis undermines the usefulness of this method in IRES studies. If not carefully ruled out by the promoterless dicistronic test, these 5'-UTRs would otherwise be considered to have IRES activities. Therefore, we propose that the promoterless dicistronic vector should be used as a control to safeguard the claim of cellular IRES in future studies.

#### ACKNOWLEDGMENTS

This work was supported in part by National Institutes of Health grants CA64539 and GM59475 and by Department of Defense grant DAMD17-02-1-0073. J.-T.Z. is a recipient of a Career Investigator Award from the American Lung Association.

We are indebted to Robert E. Rhoads for plasmid pGL2/CAT/4G/LUC, A. E. Willis for plasmids pRF and pGL3RHRV, Guntram Suske and Eddy Benveniste for plasmids pPacO and pPacSp1, and Philip Marsden for plasmid pPac-UEts-1. We also thank David Ohannesian and Chow Hwee Lee for their critical comments on the manuscript.

#### REFERENCES

- Akiri, G., D. Nahari, Y. Finkelstein, S. Y. Le, O. Elroy-Stein, and B. Z. Levi. 1998. Regulation of vascular endothelial growth factor (VEGF) expression is mediated by internal initiation of translation and alternative initiation of transcription. *Oncogene* 17:227-236.
- Bauer, C., I. Diesinger, N. Brass, H. Steinhart, H. Iro, and E. U. Meese. 2001. Translation initiation factor eIF-4G is immunogenic, overexpressed, and amplified in patients with squamous cell lung carcinoma. *Cancer* 92:822-829.
- Belsham, G. J., and N. Sonenberg. 1996. RNA-protein interactions in regulation of picornavirus RNA translation. *Microbiol. Rev.* 60:499-511.
- Bernstein, J., O. Sella, S. Y. Le, and O. Elroy-Stein. 1997. PDGF/c-sis mRNA leader contains a differentiation-linked internal ribosomal entry site (D-IRES). *J. Biol. Chem.* 272:9356-9362.
- Bonneau, A. M., and N. Sonenberg. 1987. Involvement of the 24-kDa cap-binding protein in regulation of protein synthesis in mitosis. *J. Biol. Chem.* 262:11134-11139.
- Boyer, P. L., C. Colmenares, E. Stavnezer, and S. H. Hughes. 1993. Sequence and biological activity of chicken snonN cDNA clones. *Oncogene* 8:457-466.
- Chao, D. T., and S. J. Korsmeyer. 1998. BCL-2 family: regulators of cell death. *Annu. Rev. Immunol.* 16:395-419.
- Chiang, P. W., L. E. Carpenter, and P. J. Hagerman. 2001. The 5'-untranslated region of the FMRI message facilitates translation by internal ribosome entry. *J. Biol. Chem.* 276:37916-37921.
- Clemens, M. J., and U. A. Bommer. 1999. Translational control: the cancer connection. *Int. J. Biochem. Cell Biol.* 31:1-23.
- Clemens, M. J., M. Bushell, and S. J. Morley. 1998. Degradation of eukaryotic polypeptide chain initiation factor (eIF) 4G in response to induction of apoptosis in human lymphoma cell lines. *Oncogene* 17:2921-2931.
- Coldwell, M. J., S. A. Mitchell, M. Stoneley, M. MacFarlane, and A. E. Willis. 2000. Initiation of Apaf-1 translation by internal ribosome entry. *Oncogene* 19:899-905.
- Courey, A. J., and R. Tjian. 1988. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* 55:887-898.
- Dennig, J., M. Beato, and G. Suske. 1996. An inhibitor domain in Sp3 regulates its glutamine-rich activation domains. *EMBO J.* 15:5659-5667.
- Deveraux, Q. L., and J. C. Reed. 1999. IAP family proteins—suppressors of apoptosis. *Genes Dev.* 13:239-252.
- Donahue, T. F. 2000. Genetic approaches to translation initiation in *Saccharomyces cerevisiae*, p. 487-502. In N. Sonenberg, J. W. B. Hershey, and M. B. Mathews (ed.), *Translational control of gene expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Dwarki, V. J., R. W. Malone, and I. M. Verma. 1993. Cationic liposome-mediated RNA transfection. *Methods Enzymol.* 217:644-654.
- Fen, Z., and T. O. Daniel. 1991. 5' untranslated sequences determine degradative pathway for alternate PDGF B/c-sis mRNA's. *Oncogene* 6:953-959.
- Fukuchi-Shimogori, T., I. Ishii, K. Kashiwagi, H. Mashiba, H. Ekimoto, and K. Igarashi. 1997. Malignant transformation by overproduction of translation initiation factor eIF4G. *Cancer Res.* 57:5041-5044.
- Gale, M., Jr., S. L. Tan, and M. G. Katze. 2000. Translational control of viral gene expression in eukaryotes. *Microbiol. Mol. Biol. Rev.* 64:239-280.
- Gan, W., M. LaCelle, and R. E. Rhoads. 1998. Functional characterization of the internal ribosome entry site of eIF4G mRNA. *J. Biol. Chem.* 273:5006-5012.
- Gan, W., and R. E. Rhoads. 1996. Internal initiation of translation directed by the 5'-untranslated region of the mRNA for eIF4G, a factor involved in the picornavirus-induced switch from cap-dependent to internal initiation. *J. Biol. Chem.* 271:623-626.
- Gingras, A. C., B. Raught, and N. Sonenberg. 1999. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* 68:913-963.
- Gradi, A., H. Imataka, Y. V. Svitkin, E. Rom, B. Raught, S. Morino, and N. Sonenberg. 1998. A novel functional human eukaryotic translation initiation factor 4G. *Mol. Cell. Biol.* 18:334-342.
- Hake, L. E., and J. D. Richter. 1997. Translational regulation of maternal mRNA. *Biochim. Biophys. Acta* 1332:M31-M38.
- Han, B., N. Liu, X. Yang, H. B. Sun, and Y. C. Yang. 2001. MRG1 expression in fibroblasts is regulated by Sp1/Sp3 and an Ets transcription factor. *J. Biol. Chem.* 276:7937-7942.
- Hayashi, S., K. Nishimura, T. Fukuchi-Shimogori, K. Kashiwagi, and K. Igarashi. 2000. Increase in cap- and IRES-dependent protein synthesis by overproduction of translation initiation factor eIF4G. *Biochem. Biophys. Res. Commun.* 277:117-123.
- Hellen, C. U., and P. Sarnow. 2001. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.* 15:1593-1612.
- Hentze, M. W. 1997. eIF4G: a multipurpose ribosome adapter? *Science* 275:500-501.
- Hershey, J. W. 1991. Translational control in mammalian cells. *Annu. Rev. Biochem.* 60:717-755.
- Hershey, J. W. B., and S. Miyamoto. 2000. Translational control and cancer, p. 637-654. In N. Sonenberg, J. W. B. Hershey, and M. B. Mathews (ed.), *Translational control of gene expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Holcik, M., and R. G. Korneluk. 2000. Functional characterization of the X-linked inhibitor of apoptosis (XIAP) internal ribosome entry site element: role of La autoantigen in XIAP translation. *Mol. Cell. Biol.* 20:4648-4657.
- Holcik, M., C. Lefebvre, C. Yeh, T. Chow, and R. G. Korneluk. 1999. A new internal-ribosome-entry-site motif potentiates XIAP-mediated cytoprotection. *Nat. Cell. Biol.* 1:190-192.
- Horrevoets, A. J., R. D. Fontijn, A. J. van Zonneveld, C. J. de Vries, J. W. ten Cate, and H. Pannekoek. 1999. Vascular endothelial genes that are responsive to tumor necrosis factor- $\alpha$  in vitro are expressed in atherosclerotic lesions, including inhibitor of apoptosis protein-1, stannin, and two novel genes. *Blood* 93:3418-3431.
- Horvath, P., A. Suganuma, M. Inaba, Y. B. Pan, and K. C. Gupta. 1995. Multiple elements in the 5' untranslated region down-regulate c-sis messenger RNA translation. *Cell Growth Differ.* 6:1103-1110.
- Imataka, H., A. Gradi, and N. Sonenberg. 1998. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO J.* 17:7480-7489.
- Jackson, R. J. 2000. A comparative view of initiation site selection mechanisms, p. 637-654. In N. Sonenberg, J. W. B. Hershey, and M. B. Mathews (ed.), *Translational control of gene expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Johannes, G., and P. Sarnow. 1998. Cap-independent polysomal association of natural mRNAs encoding c-myc, BIP, and eIF4G conferred by internal ribosome entry sites. *RNA* 4:1500-1513.
- Karantzoulis-Fegaras, F., H. Antoniou, S. L. Lai, G. Kulkarni, C. D'Abreo, G. K. Wong, T. L. Miller, Y. Chan, J. Atkins, Y. Wang, and P. A. Marsden. 1999. Characterization of the human endothelial nitric-oxide synthase promoter. *J. Biol. Chem.* 274:3076-3093.
- Keiper, B. D., W. Gan, and R. E. Rhoads. 1999. Protein synthesis initiation factor 4G. *Int. J. Biochem. Cell Biol.* 31:37-41.
- Khachigian, L. M., J. W. Fries, M. W. Benz, D. T. Bonthron, and T. Collins. 1994. Novel cis-acting elements in the human platelet-derived growth factor B-chain core promoter that mediate gene expression in cultured vascular endothelial cells. *J. Biol. Chem.* 269:22647-22656.
- Kozak, M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* 234:187-208.
- Kozak, M. 2001. New ways of initiating translation in eukaryotes? *Mol. Cell. Biol.* 21:1899-1907.
- Kozak, M. 1989. The scanning model for translation: an update. *J. Cell Biol.* 108:229-241.

44. Kozak, M. 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **266**:19867–19870.
45. Lamphear, B. J., R. Kirchweber, T. Skern, and R. E. Rhoads. 1995. Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. Implications for cap-dependent and cap-independent translational initiation. *J. Biol. Chem.* **270**:21975–21983.
46. Lekstrom-Himes, J., and K. G. Xanthopoulos. 1998. Biological role of the CCAAT/enhancer-binding protein family of transcription factors. *J. Biol. Chem.* **273**:28545–28548.
47. Marcu, K. B., S. A. Bossone, and A. J. Patel. 1992. myc function and regulation. *Annu. Rev. Biochem.* **61**:809–860.
48. Marcu, K. B., A. J. Patel, and Y. Yang. 1997. Differential regulation of the c-MYC P1 and P2 promoters in the absence of functional tumor suppressors: implications for mechanisms of deregulated MYC transcription. *Curr. Top. Microbiol. Immunol.* **224**:47–56.
49. Marissen, W. E., and R. E. Lloyd. 1998. Eukaryotic translation initiation factor 4G is targeted for proteolytic cleavage by caspase 3 during inhibition of translation in apoptotic cells. *Mol. Cell. Biol.* **18**:7565–7574.
50. Mathews, M. B., N. Sonenberg, and J. W. B. Hershey. 2000. Origins and principles of translational control, p. 1–32. *In* N. Sonenberg, J. W. B. Hershey, and M. B. Mathews (ed.), *Translational control of gene expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
51. McKnight, S. L. 2001. McBindall—a better name for CCAAT/enhancer binding proteins? *Cell* **107**:259–261.
52. Morris, D. R. 1995. Growth control of translation in mammalian cells. *Prog. Nucleic Acid Res. Mol. Biol.* **51**:339–363.
53. Ohlmann, T., M. Ran, V. M. Pain, and S. J. Morley. 1996. The C-terminal domain of eukaryotic protein synthesis initiation factor (eIF) 4G is sufficient to support cap-independent translation in the absence of eIF4E. *EMBO J.* **15**:1371–1382.
54. Pain, V. M. 1996. Initiation of protein synthesis in eukaryotic cells. *Eur. J. Biochem.* **236**:747–771.
55. Panniers, R. 1994. Translational control during heat shock. *Biochimie* **76**:737–747.
56. Pestova, T. V., I. N. Shatsky, and C. U. Hellen. 1996. Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Mol. Cell. Biol.* **16**:6870–6878.
57. Quandt, K., K. Frech, H. Karas, E. Wingender, and T. Werner. 1995. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* **23**:4878–4884.
58. Rao, C. D., M. Pech, K. C. Robbins, and S. A. Aaronson. 1988. The 5' untranslated sequence of the c-sis/platelet-derived growth factor 2 transcript is a potent translational inhibitor. *Mol. Cell. Biol.* **8**:284–292.
59. Ratner, L., B. Thielan, and T. Collins. 1987. Sequences of the 5' portion of the human c-sis gene: characterization of the transcriptional promoter and regulation of expression of the protein product by 5' untranslated mRNA sequences. *Nucleic Acids Res.* **15**:6017–6036.
60. Rhoads, R. E., and B. J. Lamphear. 1995. Cap-independent translation of heat shock messenger RNAs. *Curr. Top. Microbiol. Immunol.* **203**:131–153.
61. Sachs, A. B. 2000. Cell cycle-dependent translation initiation: IRES elements prevail. *Cell* **101**:243–245.
62. Sasahara, M., S. Amano, H. Sato, J. G. Yang, Y. Hayase, M. Kaneko, I. Sato, M. Suzuki, and F. Hazama. 1998. Normal developing rat brain expresses a platelet-derived growth factor B chain (c-sis) mRNA truncated at the 5' end. *Oncogene* **16**:1571–1578.
63. Schneider, R., and M. Kozak. 2001. New ways of initiating translation in eukaryotes? *Mol. Cell. Biol.* **21**:8238–8246.
64. Schneider, R. J. 2000. Translational control during heat shock, p. 581–593. *In* N. Sonenberg, J. W. B. Hershey, and M. B. Mathews (ed.), *Translational control of gene expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
65. Sella, O., G. Gerlitz, S. Y. Le, and O. Elroy-Stein. 1999. Differentiation-induced internal translation of c-sis mRNA: analysis of the cis elements and their differentiation-linked binding to the hnRNP C protein. *Mol. Cell. Biol.* **19**:5429–5440.
66. Sharrocks, A. D., A. L. Brown, Y. Ling, and P. R. Yates. 1997. The ETS-domain transcription factor family. *Int. J. Biochem. Cell Biol.* **29**:1371–1387.
67. Shinagawa, T., H. D. Dong, M. Xu, T. Maekawa, and S. Ishii. 2000. The sno gene, which encodes a component of the histone deacetylase complex, acts as a tumor suppressor in mice. *EMBO J.* **19**:2280–2291.
68. Sonenberg, N. 1994. mRNA translation: influence of the 5' and 3' untranslated regions. *Curr. Opin. Genet. Dev.* **4**:310–315.
69. Sonenberg, N. 1994. Regulation of translation and cell growth by eIF-4E. *Biochimie* **76**:839–846.
70. Stoneley, M., T. Subkhankulova, J. P. Le Quesne, M. J. Coldwell, C. L. Jopling, G. J. Belsham, and A. E. Willis. 2000. Analysis of the c-myc IRES; a potential role for cell-type specific trans-acting factors and the nuclear compartment. *Nucleic Acids Res.* **28**:687–694.
71. Vagner, S., B. Galy, and S. Pyronnet. 2001. Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites. *EMBO Rep.* **2**:893–898.
72. Wasyluk, B., S. L. Hahn, and A. Giovane. 1993. The Ets family of transcription factors. *Eur. J. Biochem.* **211**:7–18.
73. Willis, A. E. 1999. Translational control of growth factor and proto-oncogene expression. *Int. J. Biochem. Cell Biol.* **31**:73–86.
74. Yan, R., W. Rychlik, D. Etchison, and R. E. Rhoads. 1992. Amino acid sequence of the human protein synthesis initiation factor eIF-4 gamma. *J. Biol. Chem.* **267**:23226–23231.
75. Zhang, J. T., and V. Ling. 1991. Study of membrane orientation and glycosylated extracellular loops of mouse P-glycoprotein by in vitro translation. *J. Biol. Chem.* **266**:18224–18232.



## Regulation of constitutive expression of mouse PTEN by the 5'-untranslated region

Baoguang Han<sup>1</sup>, Zizheng Dong<sup>1</sup>, Yang Liu<sup>1</sup>, Qun Chen<sup>1</sup>, Katsuyuki Hashimoto<sup>2</sup> and Jian-Ting Zhang<sup>\*,1</sup>

<sup>1</sup>Department of Pharmacology and Toxicology, I.U. Cancer Center and Walther Oncology Center/Walther Cancer Institute, Indiana University School of Medicine, Indianapolis, IN 46202, USA; <sup>2</sup>National Institute of Infectious Diseases, Tokyo, Japan

PTEN tumor suppressor serves as a major negative regulator of survival signaling mediated by PI3 kinase/AKT/protein kinase B pathway, and is inactivated in various human tumors. Elucidation of mechanisms responsible for PTEN expression is important for providing insight into strategies to control the loss of PTEN expression in human cancers. Although recent studies suggested that p53 and Egr-1 can modulate induced PTEN expression, the mechanism responsible for ubiquitous constitutive expression of PTEN remains elusive. PTEN mRNA contains a highly conserved and GC-rich 5'-untranslated region (5'-UTR). Recently, it has been shown that the long 5'-UTR sequences of several growth-regulated mRNAs contain promoters that can generate mRNAs with shorter 5'-UTRs. In this paper, we tested whether the 5'-UTR sequence of mouse PTEN contains a promoter that is responsible for constitutive expression of PTEN. We found that the long 5'-UTR sequence of mouse PTEN severely inhibits translation of PTEN and a heterologous gene firefly luciferase. Deletion of the most 5'-UTR sequence would enhance translation efficiency 100-fold. We also showed that the 5'-UTR sequence of mouse PTEN does not have an internal ribosome entry site (IRES) that can mediate cap-independent initiation of translation. Instead, we found that the 5'-UTR sequence of mouse PTEN contains a strong promoter that drives the production of a transcript with shorter 5'-UTRs, which can be translated with higher efficiency. This promoter was mapped to the region between –551 and –220 bases upstream of the translation start codon. Cotransfection analysis using *Drosophila* SL2 cells showed that Sp1 is one of the major transcription factors that can constitutively activate this promoter. Two endogenous PTEN transcripts with 5'-UTRs of 193 and 109 bases were found in DU145 and H226 cell lines. Based on these observations, we conclude that the PTEN expression may be regulated at both transcriptional and translational levels, and that the 5'-UTR sequence of PTEN contains a

promoter that is responsible for constitutive PTEN expression.

*Oncogene* (2003) 22, 5325–5337. doi:10.1038/sj.onc.1206783

**Keywords:** PTEN; IRES; promoter; 5'-UTR

### Introduction

PTEN/MMAC1/TEP1 (referred as PTEN in the remaining text of this paper) is a tumor suppressor gene that maps to the 10q23.3 and has been shown to be deleted or mutated in many human tumors including glioblastomas, endometrial neoplasms, hematological malignancies, and prostate and breast cancers (Ali *et al.*, 1999; Cantley and Neel, 1999; Dahia, 2000; Di Cristofano and Pandolfi, 2000). In addition, germline mutations in PTEN cause Cowden syndrome (CS), characterized by multiple hamartomas and a high proclivity for developing cancer (Cantley and Neel, 1999). Although PTEN sequence is highly homologous with dual-specificity protein phosphatase, its major substrate is phosphatidylinositol triphosphate (PIP<sub>3</sub>). It has been suggested that PTEN is a major negative regulator of PIP<sub>3</sub> and negatively regulates the survival signaling mediated by the PI3 kinase/AKT/protein kinase B pathway (Maehama and Dixon, 1999; Leslie and Downes, 2002). Loss of PTEN function or expression results in an increased concentration of PIP<sub>3</sub> and AKT hyperactivation, which leads to the protection of cells from various apoptotic stimuli (Stambolic *et al.*, 1998). In contrast, overproduction of PTEN induces growth suppression via cell cycle arrest and/or induction of apoptosis, and inhibits cell adhesion and migration (Dahia, 2000). Recently, PTEN has been demonstrated to serve as a negative regulator for proliferation of human neural stem cells and mammary epithelial cells (Backman *et al.*, 2001; Groszer *et al.*, 2001; Kwon *et al.*, 2001; Li *et al.*, 2002).

With the critical role in antagonizing PI3 kinase pathways, PTEN is anticipated to be the target of complex control mechanisms (Leslie and Downes, 2002). However, very little is known about the regulation of PTEN expression. PTEN expression is ubiquitous and constitutive, while it can also be altered by

\*Correspondence: Jian-Ting Zhang; Department of Pharmacology and Toxicology, IUCC, Indiana University School of Medicine, 1044 W. Walnut Street, R4-166, Indianapolis, IN 46202, USA;  
E-mail: jianzhan@iupui.edu

Received 15 November 2002; revised 13 May 2003; accepted 13 May 2003

various biological stimuli. It was reported that PTEN expression in human keratinocytes decreased when the cells were treated with TGF- $\beta$  (Li and Sun, 1997) and PTEN expression increased dramatically when myeloid leukemia cells were induced to differentiate into either granulocytic or monocytic cells (Hisatake *et al.*, 2001). PTEN expression is also induced during neuronal differentiation (Ross *et al.*, 2001). It has been reported that PTEN expression can be induced by p53 and Egr-1, and functional p53 and Egr-1 binding sites were identified between bases -1190 and -1157, and between -947 and -939 upstream of the translation start codon (Stambolic *et al.*, 2001; Virolle *et al.*, 2001), respectively. This observation suggests that p53 and Egr-1 may be able to modulate the PI3 kinase/AKT/Protein kinase B pathway via regulating PTEN expression (Stambolic *et al.*, 2001). However, mechanisms responsible for constitutive expression of PTEN remain to be identified.

Recently, we proposed that promoters may exist in the long 5'-UTR sequences of genes related to cell growth (Han and Zhang, 2002). In this study, we attempted to address this hypothesis by studying the regulation mechanism of PTEN expression. Examination of the 5' untranslated region (5'-UTR) of PTEN messenger RNA (mRNA) revealed several interesting features related to translational regulation. First, the 5'-UTR of PTEN mRNA is about 1 kb inferred from cDNA sequences of both mouse (GeneBank Accession number NM\_008960) and human (GeneBank Accession number NM\_000314) PTEN and confirmed by reverse transcription-polymerase chain reaction (RT-PCR) (Stambolic *et al.*, 2001). This length of 5'-UTR is significantly longer than the average length of 100–300 bases for eucaryotic 5'-UTRs. Second, the 5'-UTR of both mouse and human PTEN mRNA contains four upstream translation start codons. Third, the 5'-UTR of PTEN mRNA is G/C-rich with a predicted complex secondary structure. Lastly, the 5'-UTRs of both human and mouse PTEN mRNA are well conserved with an overall homology of >90%. These features of PTEN mRNA strongly suggest that the 5'-UTR of PTEN may inhibit efficient translation initiation by the conventional cap-dependent ribosome scanning mechanism. It is possible that PTEN mRNAs use other mechanisms such as internal ribosome entry site (IRES) to initiate translation (for reviews see Pestova *et al.* (2001) and Sachs (2000)). Alternatively, the long 5'-UTR of PTEN may contain a promoter or splice site to generate mRNA species with significantly shorter 5'-UTRs (Kozak, 2001; Schneider and Kozak, 2001), which can be translated efficiently (Han and Zhang, 2002).

In this study, we tested these possibilities and found that mRNAs with long 5'-UTRs of PTEN cannot be translated by the cap-dependent mechanism, whereas the mRNAs with shorter 5'-UTRs can. The long 5'-UTR of PTEN does not contain an IRES element to mediate IRES-dependent translation initiation. Instead, we demonstrated that the 5'-UTR of PTEN contains a strong promoter that produces mRNA with shorter 5'-UTRs. Using 5' RACE and RNase protection assay, we found two endogenous PTEN transcripts with short

5'-UTRs of 109 and 193 bases, respectively, in human cancer cell lines. These mRNA species with shorter 5'-UTRs may be responsible for constitutive production of PTEN proteins. Interestingly, the promoter region in the 5'-UTR overlaps with the hypermethylated region identified in gastric tumor (Kang *et al.*, 2002) and endometrial carcinoma (Salvesen *et al.*, 2001), implicating that hypermethylation of this promoter may be important for the loss of PTEN expression observed in a number of tumors.

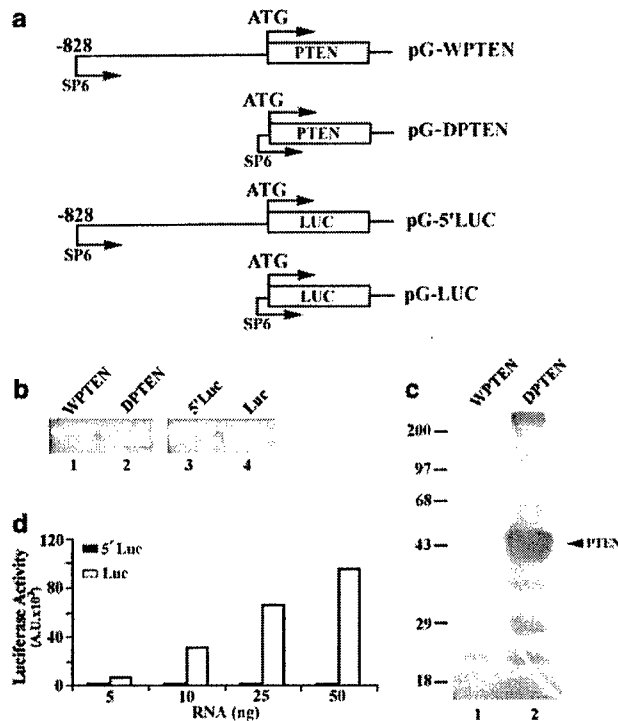
## Results

### *The Long 5'-UTR of mouse PTEN mRNA inhibits translation in vitro*

The 5'-UTR of mouse PTEN mRNA is 948-nucleotide long and contains 72% G and C residues. Analysis of the 5'-UTR using Zuker's mfold program (version 3.0) (Mathews *et al.*, 1999) showed that this region has potential to form various secondary structures and all of the structures predicted are extremely stable with a typical free energy as high as -432 kcal/mol, which is well beyond the energy shown to inhibit ribosome scanning. In addition, this region is also punctuated with four translation start codon AUGs that encode two short upstream open reading frames (uORFs) with 36 and 45 amino acids, respectively (two of the AUGs are located at the end of the first uORF in the same frame). Both uORFs end before -551 bases upstream of the physiological translation start codon. Generally, uAUGs or uORFs inhibit translation of the main ORF by interfering with the scanning process of the 40S ribosome complex. To analyse the effect of the 5'-UTR sequence of mouse PTEN on translation, we utilized an *in vitro* transcription and translation system. Mouse PTEN cDNA with (pG-WPTEN) and without (pG-DPTEN) the 5'-UTR sequence (Figure 1a) were cloned into pGEM-4Z under the control of an SP6 promoter. Run-off transcripts were generated by SP6 RNA polymerase, purified and quantified (Figure 1b), and an equal amount of RNAs were used to program cell-free translation in rabbit reticulocyte lysate (RRL). As shown in Figure 1c, no full-length PTEN protein was translated from transcripts containing the long 5'-UTR (WPTEN) (lane 1), whereas a significant amount of PTEN protein (DPTEN) was generated in the absence of the 5'-UTR sequence (lane 2). Clearly, the presence of the long 5'-UTR sequence drastically inhibits the translation of PTEN in RRL.

To analyse whether the protein-encoding sequence (ORF) influences the translational inhibition effect of the 5'-UTR, we cloned the 5'-UTR sequence of PTEN immediately upstream of an open reading frame encoding firefly luciferase (LUC) in a pGEM-4Z-based reporter vector, pG-LUC, resulting in the plasmid pG-5'LUC (Figure 1a). This construct was then used to program *in vitro* transcription (Figure 1b) and cell-free translation (Figure 1d) as described above. Again, the translation programmed by the pG-LUC transcript

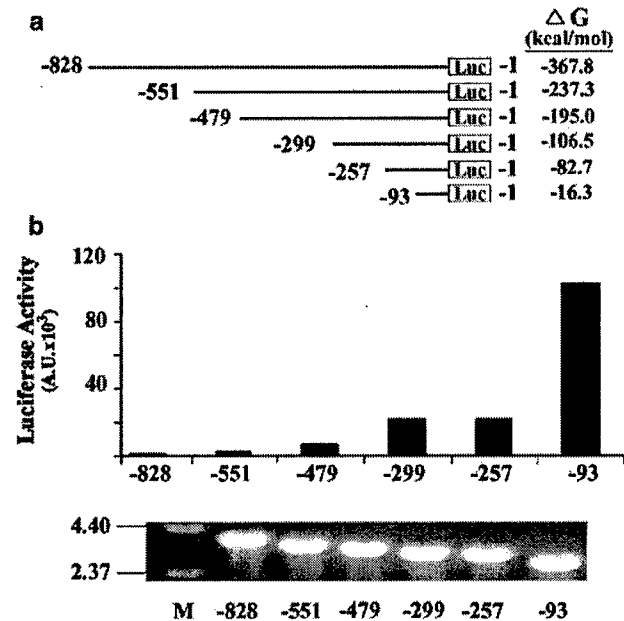




**Figure 1** 5'-UTR of PTEN inhibits translation of its downstream sequence in RRL. (a) Schematic diagram of *in vitro* transcription constructs. The transcription start site using SP6 RNA polymerase and the translation start codon AUG are indicated by arrows. (b) *In vitro* transcription. The capped RNA transcripts were generated by *in vitro* transcription using linearized plasmids and SP6 RNA polymerase, and 500 ng of each RNA transcript was separated using 1% agarose gel. (c) *In vitro* translation of mouse PTEN. The *in vitro* transcripts of mouse PTEN with (WPTEN) or without (DPTEN) the 5'-UTR sequence were used to program translation in RRL in the presence of [<sup>35</sup>S]methionine. Protein product was separated by SDS-PAGE and visualized by autoradiography. (d) *In vitro* translation of firefly luciferase. 5, 10, 25, and 50 ng of *in vitro* transcripts of firefly luciferase with (5'LUC, solid bars) or without (LUC, open bars) the 5'-UTR sequence of mouse PTEN were used to program translation in RRL, followed by measuring luciferase activity by enzymatic assays

generated a significantly increasing amount of firefly LUC activity proportional to the amount of transcripts used, while no LUC activity was detected from the translation programmed by pG-5'LUC transcripts. Thus, the 5'-UTR sequence of mouse PTEN inhibits translation of transcripts encoding a heterologous protein and is independent of the protein-encoding sequence.

To further analyse the translational inhibition effect of the 5'-UTR sequence of PTEN, we generated several sequential deletion mutants from the 5' end of the 5'-UTR in pG-5'LUC constructs (Figure 2a). These deletion mutants contain putative secondary structures with different free energies from -367 to -16 kcal/mol predicted by Zuker's mfold program (Mathews *et al.*, 1999). As shown in Figure 2b, translation of firefly LUC was very inefficient for transcripts with 5'-UTRs that have 479 nucleotides or longer. Further deletion resulted in an increase in translation. However, significant translation occurred only to the transcript with 93 bases of the 5'-UTR sequence with a predicted energy of



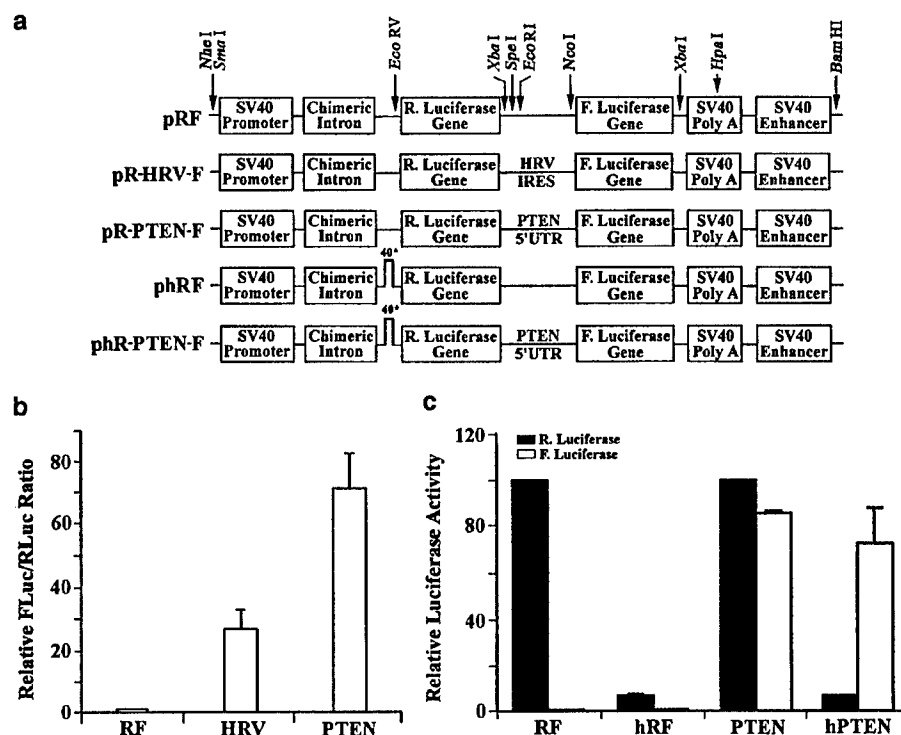
**Figure 2** Effect of deletions of the 5'-UTR sequence of mouse PTEN on the translation of firefly luciferase in RRL. (a) Schematic diagram of sequential deletions in the 5'-UTR sequence of mouse PTEN. The free energy for the predicted secondary structure of the 5'-UTR sequence is shown on the right. The positions of the 5' end of each deletion are indicated on the left. These mutant 5'-UTR sequences were engineered into pG-LUC for *in vitro* expression. (b) *In vitro* transcription and translation. The deletion mutants were used to generate *in vitro* transcripts (bottom), which were then used to program cell-free translation in RRL. LUC activities of the translated products were measured by enzymatic assays as described in Materials and methods

-16 kcal/mol. It is ~100-fold more translatable than the full-length transcript. These results are consistent with a model that the stable secondary structure in the 5'-UTR sequence of PTEN mRNA inhibits the ribosome scanning process for cap-dependent translation initiation.

#### 5'-UTR sequence of PTEN enhances expression of the second cistron in dicistronic test

The above results clearly demonstrate that the 5'-UTR of PTEN mRNA can effectively suppress ribosome scanning *in vitro*. However, PTEN protein has been reported to be constitutively expressed and its expression can be induced to a high level by biological stimulation. Thus, if PTEN mRNA with the full-length 5'-UTR is translated, it may use an alternative mechanism such as IRES-mediated translation initiation. To test this hypothesis, the 5'-UTR (-828 to -1 bases) of PTEN was cloned into the intergenic region of a dicistronic vector pRF (Stoneley *et al.*, 2000) to obtain pR-PTEN-F (Figure 3a). The internal ribosome entry site (IRES) sequence of human rhinovirus (HRV) was engineered in the same way and was used as a positive control. The pRF-based constructs contain an SV40 promoter to direct the transcription of dicistronic RNA encoding *Renilla* LUC as the first cistron and firefly LUC as the second cistron. Translation of the first





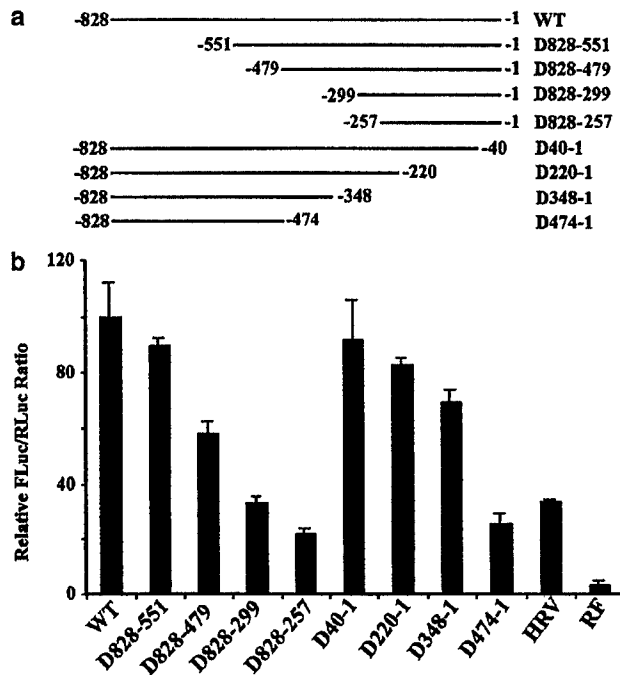
**Figure 3** Dicistronic DNA test of the 5'-UTR sequence of mouse PTEN. (a) Schematic diagram of dicistronic constructs. The 5'-UTR sequence of mouse PTEN and the IRES element of HRV are cloned into the intergenic region. The inverted U-shaped structure in phRF and phR-PTEN-F represents the inserted hairpin loop with a free energy of  $-40$  kcal/mol. The locations of several relevant restriction enzyme sites are shown by arrows. (b) Relative LUC activities generated by the dicistronic constructs in HeLa cells. HeLa cells were transfected with pRF, pR-HRV-F, and pR-PTEN-F constructs. At 24 h following transfection, cells were harvested and the *Renilla* and firefly LUC activities were measured and the relative ratios were calculated and normalized to that of the vector-transfected cells (RF). (c) Effect of hairpin structure on the expression of LUCs. HeLa cells were transfected with phRF and phR-PTEN-F, which has a hairpin structure immediately upstream of the ATG codon of the *Renilla* LUC gene. At 24 h following transfection, cells were harvested and *Renilla* and firefly LUC activities were measured in comparison with that generated by pRF and pR-PTEN-F, respectively. Both *Renilla* (solid bars) and firefly LUC (open bars) activities were normalized to pRF control. All the data were from three independent experiments

cistron (*Renilla* LUC) serves as an indicator of cap-dependent translation, while translation of the second cistron (firefly LUC) reflects the IRES activity-associated with the inserted intergenic sequence. This approach is considered as a 'gold standard' for characterizing cellular IRES (Sachs, 2000). These dicistronic constructs were transfected into HeLa cells and both *Renilla* and firefly LUC activities were measured. As shown in Figure 3b, the 5'-UTR sequence of PTEN stimulated the expression of firefly LUC by  $\sim 70$ -fold over negative vector control and about 2–3-fold over positive pR-HRV-F control. Similar results were also obtained using another cell line H1299 (data not shown). These results suggest that the 5'-UTR sequence of PTEN may contain an IRES element or it may enhance read-through or jumping from the first cistron.

To determine whether the effect of the 5'-UTR sequence of PTEN on the translation of the second cistron is due to enhanced read-through or jumping from the first cistron, we inserted a synthetic hairpin in front of the first cistron and created phRF and phR-PTEN-F (Figure 3a). The hairpin structure has a free energy of  $-40$  kcal/mol and is expected to significantly inhibit the cap-dependent translation initiation of the

first cistron. Therefore, it will decrease potential ribosome read-through or jumping but will not affect the translation from IRES-mediated initiation of the second cistron. As shown in Figure 3c, the insertion of the hairpin resulted in  $\sim 90\%$  decrease of the *Renilla* LUC activity (first cistron) in both phRF and phR-PTEN-F as compared with pRF and pR-PTEN-F, respectively, suggesting that the hairpin structure inhibits the cap-dependent translation. However, the level of firefly LUC activity (second cistron) produced by phR-PTEN-F was not significantly affected when compared with pR-PTEN-F. This result suggests that the enhanced expression of the second cistron by the 5'-UTR sequence of PTEN was independent of the cap-dependent translation initiation of the first cistron. Thus, it is not due to enhanced read-through or jumping from the first cistron.

To define the boundaries in the 5'-UTR sequence of PTEN for enhancing expression of the second cistron, a series of deletion mutants of the 5'-UTR of PTEN were engineered for a dicistronic assay in HeLa cells (Figure 4a). As shown in Figure 4b, deletion from both ends of the 5'-UTR resulted in a gradual decrease in stimulating firefly LUC expression. A construct pR-hPTEN-F that

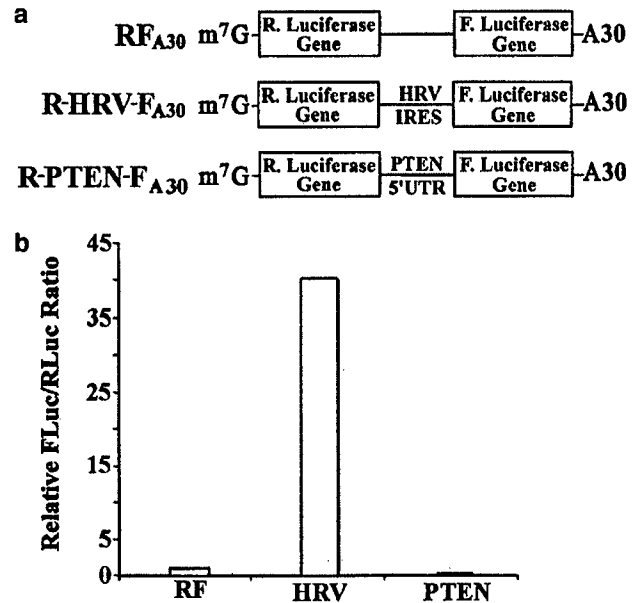


**Figure 4** Deletion mapping of the 5'-UTR sequence of mouse PTEN in dicistronic DNA test. (a) Schematic diagram of sequential deletions of the 5'-UTR sequence of mouse PTEN. The positions of the 5' and 3' ends of each deletion are indicated on the left and right, respectively. These mutant 5'-UTR sequences were engineered into the dicistronic vector pRF at the intergenic region. (b) Relative luciferase activities generated from constructs containing the wild-type and mutant 5'-UTR sequences of mouse PTEN. HeLa cells were transfected with the constructs shown in (a), and 24 h following transfection, *Renilla* and firefly LUC activities were measured and the ratio of firefly to *Renilla* LUC was calculated and normalized to the wild-type (WT) control. The empty vector pRF and the plasmid containing HRV IRES were also used as negative and positive controls, respectively. The data were from three independent experiments

encodes the -492 to -1 region of the 5'-UTR sequence of human PTEN enhanced the expression of the second cistron by 50-fold (data not shown), equivalent to that of the D828-479 construct of mouse PTEN (Figure 4b). Similar results with these constructs have also been observed by using another cell line H1299 (data not shown). Since the deletion mutant D828-551 and D220-1 retains the most enhancing activity, the 230-nucleotide central region of the 5'-UTR (-551 to -220) may contain important elements for stimulating firefly LUC expression. It is noteworthy that deletion mutants that retain about 200 nucleotides of the 5'-UTR sequence at the 3'-end (D828-257) or at the 5'-end (D474-1) still have the enhancing activity comparable to HRV IRES, suggesting that the enhancing activity also exists at both 5'- and 3'-ends of the 5'-UTR sequence of PTEN.

#### 5'-UTR of PTEN does not display an IRES activity in dicistronic mRNA assay

The above results suggest that the 5'-UTR of PTEN may (1) contain an IRES activity that enhances the translation of firefly LUC from the dicistronic mRNA



**Figure 5** Translation of dicistronic mRNA in HeLa cells. (a) Schematic diagram of the dicistronic mRNA used for translation in HeLa cells. *In vitro* transcripts with 5' cap (m7G) and 3' poly A tail (A<sub>30</sub>) were synthesized using T7 RNA polymerase from linearized vector (RF<sub>A30</sub>) and constructs containing the IRES of HRV (R-HRV-F<sub>A30</sub>) and the 5'-UTR of PTEN (-551 to -1) (R-PTEN-F<sub>A30</sub>). (b) Relative LUC activity from dicistronic mRNAs in HeLa cells. HeLa cells were transfected with the dicistronic mRNAs, and 8 h following transfection, *Renilla* and firefly LUC activities were measured and the relative ratios were calculated and normalized to that of the vector-transfected cells (RF<sub>A30</sub>)

by internal initiation (for reviews see Pestova *et al.* (2001) and Sachs (2000)), (2) contain a promoter that directs transcription of the firefly LUC gene (Han and Zhang, 2002), and/or (3) contain a splicing acceptor site, that creates a splicing variant with only the second cistron of the firefly LUC gene. To distinguish between these possibilities, we generated dicistronic RNAs *in vitro* from the dicistronic constructs and used them to program translation both in HeLa cells and in RRL. RNA transfection allows a direct analysis of whether the 5'-UTR sequence of PTEN in the intergenic region of a dicistronic mRNA can enhance the translation of the second cistron without transcriptional interference. For purposes of RNA transfection, plasmids pSP-RF<sub>A30</sub>, pSP-R-PTEN-F<sub>A30</sub>, and pSP-R-HRV-F<sub>A30</sub> were engineered and used for producing dicistronic transcripts containing m7GpppG cap and polyadenylated tail *in vitro* (Figure 5a). The *in vitro* transcripts were introduced into HeLa cells by lipofectin encapsulation. At 8 h following transfection, cell lysates were prepared for LUC activity measurement. As expected, the firefly LUC of vector RNA was very poorly translated and its activity (arbitrary units) only represented about 0.16% of *Renilla* LUC (data not shown). It increased to about 6.7% with the dicistronic RNAs containing HRV IRES. Therefore, HRV IRES significantly stimulated the translation of firefly LUC about 40-fold over vector control (Figure 5b). However, no stimulation of firefly LUC expression was observed with the 5'-UTR of

PTEN. The expression of the second cistron firefly LUC in the presence of the 5'-UTR of PTEN was about 0.003% of *Renilla* LUC activity, significantly less than that observed with vector control. Again, this observation confirms the conclusion that the presence of the 5'-UTR sequence of PTEN significantly prevents ribosome scanning initiated from the 5' end. Similar results were observed using the dicistronic RNA transcripts to program cell-free translation in RRL supplemented with HeLa extract (data not shown). Thus, it appears that the 5'-UTR sequence of PTEN does not contain an IRES element to mediate internal ribosome entry.

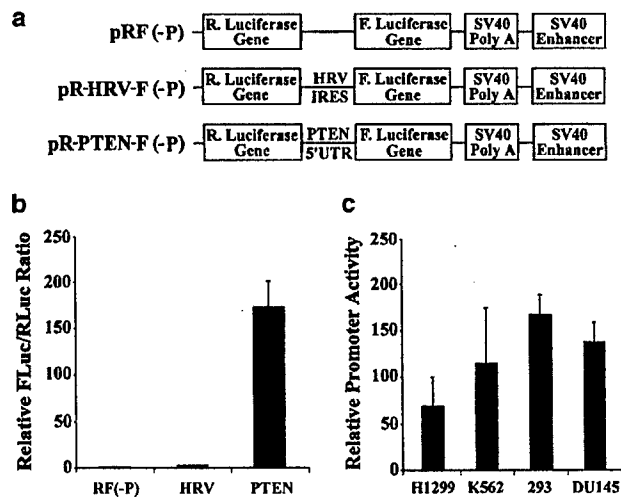
#### 5'-UTR sequence of PTEN contains a ubiquitously functional promoter

The above results prompted us to explore whether the 5'-UTR sequence of PTEN contains a promoter. For this purpose, we used a promoterless dicistronic assay as described previously (Han and Zhang, 2002) by simply removing the unique SV40 promoter together with the intron sequence from the pRF-based dicistronic constructs. These promoterless dicistronic constructs (Figure 6a) were then transfected into HeLa cells for determination of both *Renilla* and firefly LUC activities (Figure 6b). As we have shown previously (Han and

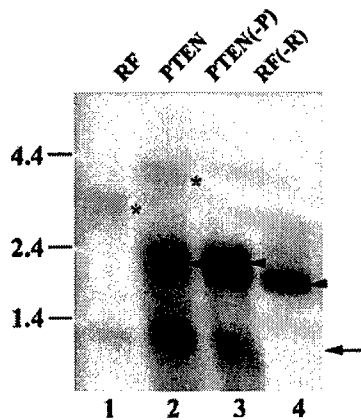
Zhang, 2002), both the *Renilla* and firefly LUC activities were minimal, but detectable for pRF(-P) vector control. Only a twofold increase in firefly LUC activity was observed with pR-HRV-F(-P) construct. This small increase was in dramatic contrast with the 30-fold increase associated with the pR-HRV-F construct (Figure 3b). Thus, the enhanced expression of firefly LUC from pR-HRV-F construct was not due to production of monocistronic transcripts by a promoter present in the HRV sequence. However, the pR-PTEN-F(-P) construct generated more than 160-fold firefly LUC activity over vector control. Thus, the significant stimulation of the firefly LUC expression by the 5'-UTR sequence of PTEN was likely due to the presence of a strong promoter in this region.

The existence of a constitutively active promoter in the 5'-UTR sequence of PTEN is unexpected and may help explain the relatively constant and high levels of PTEN mRNA and protein in most primary cells (Stambolic *et al.*, 2001). To determine whether this promoter is ubiquitously used, we analysed its activity in various human cell lines including HEK293 (transformed primary embryonic kidney cells), H1299 (large-cell lung carcinoma cells), K562 (erythroleukemia cells), and DU145 (prostate cancer cells). As shown in Figure 6c, the 5'-UTR sequence displays significant promoter activities in all the cell lines tested, at a level similar to that observed in HeLa cells. Thus, this promoter in the 5'-UTR of PTEN is ubiquitously and constitutively active.

To determine directly whether the transcript derived from the promoter in the 5'-UTR sequence of PTEN exists in cells, poly(A) mRNAs were isolated for Northern blot analysis 24 h following transfection of pRF, pR-PTEN-F, pR-PTEN-F(-P), and pRF(-R). The pRF(-R) construct, which lacks the *Renilla* LUC gene (Han and Zhang, 2002), was used as a monocistronic control. This vector is expected to produce transcript encoding only firefly LUC using the SV40 promoter. The pRF vector was used as a dicistronic control, which produces only dicistronic mRNA using SV40 promoter. As shown in Figure 7, dicistronic transcript from control pRF (lane 1, indicated by an asterisk) and the monocistronic transcript from control pRF(-R) (lane 4, indicated by an arrowhead) were detected as expected. The dicistronic transcript from pR-PTEN-F (lane 2, indicated by an asterisk) is bigger than that from pRF (lane 1), consistent with the presence of the 5'-UTR sequence of PTEN in the intergenic region. The transcript produced from pR-PTEN-F(-P) has a size slightly larger than that from pRF(-R) (compare lanes 3 and 4), suggesting that it is a monocistronic mRNA produced by the promoter in the 5'-UTR of PTEN. The same monocistronic transcript was also observed with pR-PTEN-F (lane 2, indicated by an arrowhead). These observations confirm that the 5'-UTR sequence of PTEN has a strong promoter for transcription of the second cistron firefly LUC. It is noteworthy that the monocistronic RNA bands from both pR-PTEN-F and pR-PTEN-F(-P) on Northern blot (lanes 2 and 3, indicated by an arrowhead) are much wider than that



**Figure 6** 5'-UTR sequence of mouse PTEN contains a ubiquitously active promoter. (a) Schematic diagram of promoterless dicistronic construct of pRF(-P), pR-HRV-F(-P), and pR-PTEN-F(-P). The sequences of SV40 promoter and chimeric intron were removed from the parental dicistronic constructs shown in Figure 3. (b) Relative LUC activities generated from the promoterless dicistronic constructs. The promoterless constructs were transfected into HeLa cells, and 24 h following transfection, cells were harvested for determination of *Renilla* and firefly luciferase activity. The relative ratios between firefly and *Renilla* LUC activities were calculated and normalized to that of the vector-transfected cells (pRF(-P)). (c) Promoter activity of the 5'-UTR sequence in various human cell lines. Human H1299, K562, HEK293, and DU145 cells were transfected with promoterless constructs and cell lysates were prepared 24 h following transfection for LUC measurement as described for HeLa cells in panel b. Relative promoter activity was calculated as the ratio of firefly LUC activity of pR-PTEN-F(-P) to the vector control pRF(-P). The data were from three independent experiments



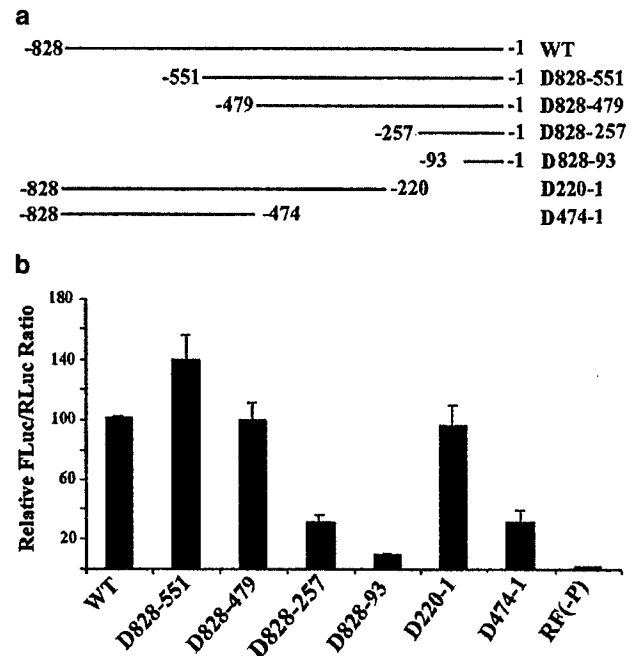
**Figure 7** Northern blot analysis of RNA products generated by promoter of mouse PTEN. PolyA mRNAs were isolated following transfection with pRF (lane 1), pR-PTEN-F (lane 2), pR-PTEN-F(-P) (lane 3), and pRF(-R) (lane 4) and used for Northern blot analysis as described in Materials and methods. The asterisks and arrowheads indicate the dicistronic and monocistronic RNA transcripts, respectively. The arrow shows unknown RNAs hybridized with the probe which have been consistently observed in previous studies (Coldwell *et al.*, 2000; Han and Zhang, 2002)

from pRF(-R) (lane 4), suggesting that these transcripts may be derived from multiple transcription start sites. Indeed, RT-PCR analysis showed the existence of transcripts starting from a very upstream region of the 5'-UTR (data not shown).

#### Transactivation of the promoter in the 5'-UTR sequence of PTEN by Sp1

To identify the boundaries of the DNA region that are responsible for promoter activity, deletion mutants were generated from either 5' or 3' ends of the 5'-UTR of PTEN (Figure 8a). These deletion mutants were engineered into the promoterless dicistronic vector and then used to determine their ability to direct firefly LUC expression. As shown in Figure 8b, deletion from either end of the 5'-UTR resulted in a gradual decrease in stimulating firefly LUC expression. The full stimulation activity resides in the region of -551 to -220 since deletion mutants D828-551 and D220-1 still contain full-promoter activity. In addition, the deletion mutants containing only the 5' (D474-1) or 3' (D828-257) region of the 5'-UTR sequence also have significant promoter activity. Widespread promoter activity within the 5'-UTR sequence of PTEN may be responsible for multiple transcriptional initiation sites observed with Northern blots. These observations are consistent with the results obtained from conventional dicistronic DNA constructs (see Figure 4).

Analysis of the potential transcription factor binding sites in the -551 to -220 region of the 5'-UTR sequence using MatInspector (Quandt *et al.*, 1995) showed several consensus sites for transcription factors such as Sp1, Ets-1, Egr-1, Ap-1, Ap-2, and Ap-4 (Figure 9a). To determine directly whether the above transcription factors could functionally modulate promoter activity in the 5'-UTR sequence of PTEN, *Drosophila* SL2 cells,

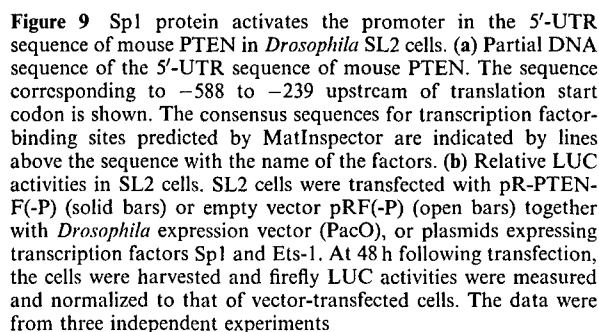


**Figure 8** Deletion mapping of the promoter activity in the 5'-UTR sequence of mouse PTEN. (a) Schematic diagram of the deletions in the 5'-UTR of mouse PTEN. The positions of 5' and 3' ends of each deletion are indicated on the left and right, respectively. These mutant 5'-UTRs were engineered into the promoterless dicistronic vector pRF(-P) at the intergenic region. (b) Relative LUC activity generated from the wild-type and mutant 5'-UTR sequences of mouse PTEN. HeLa cells were transfected with the constructs shown in (a), and 24 h following transfection, *Renilla* and firefly LUC activities were measured and the ratio of firefly to *Renilla* LUC was calculated and normalized to the wild-type (WT) control. The data were from three independent experiments

which are deficient in Sp1-, Sp3-, and Ets-related proteins (Courey and Tjian, 1988; Dennig *et al.*, 1996), were used. The reason for using insect instead of mammalian cells is that Sp1-, Sp3-, and Ets-related factors are expressed in virtually all mammalian cells, which could affect the interpretation of this experiment. We introduced the pR-PTEN-F(-P) construct along with *Drosophila* expression plasmids pPacSp1 or pPacUets-1 into *Drosophila* SL2 cells. As shown in Figure 9b, pPacSp1 stimulated the promoter activity in the 5'-UTR sequence of PTEN by 15-fold, whereas it stimulated vector control only by threefold. Ets-1 did not significantly stimulate the promoter activity in the 5'-UTR sequence of PTEN. These results suggest that Sp1 is a possible transcription factor that regulates the constitutive expression of PTEN by acting at the promoter within the 5'-UTR of PTEN. Since there are several Sp1 binding sites in the 5'-UTR sequence, it remains to be determined which Sp1 element is responsible for the promoter activity.

#### Detection of endogenous transcripts originating from the 5'-UTR promoter in human cancer cell lines

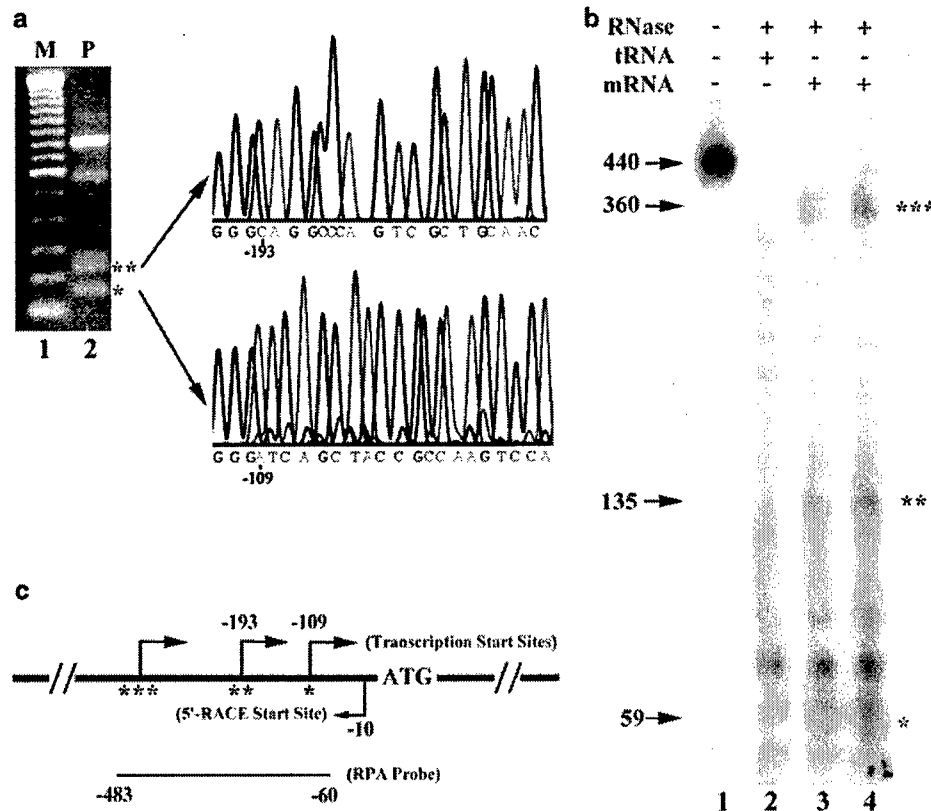
Previous research with human PTEN using Northern blot detected heterogeneous transcripts of a variety of



As an independent approach to demonstrate the existence of endogenous PTEN transcripts with shorter 5'-UTRs, we performed an RNase protection assay. Using a probe covering the region of -483 to -60 in the 5'-UTR sequence of human PTEN (Figure 10c), we generated three protected products with estimated sizes of 360, 135, and 59 bases, respectively, with RNAs isolated from both DU145 and H226 cell lines (Figure 10b). However, no protected RNAs were found with yeast tRNA control. While the products of 135 and 59 base bands are likely derived from transcripts with 5'-UTRs of 109 and 193 bases as shown by 5' RACE, respectively, the product of 360 bases is possibly produced from the transcripts with a longer 5'-UTR. Together, these results demonstrate that endogenous PTEN transcripts with shorter 5'-UTRs exist in human cell lines and they are likely responsible for efficient translation to produce proteins.

## Discussion

The long and structured 5'-UTR sequence of PTEN mRNA is highly conserved between mouse and human. Conservation of the 5'-UTR sequence is also observed for other proto-oncogenes or factors such as VEGF and PDGF (Bernstein *et al.*, 1997; Huez *et al.*, 1998). The conservation of the 5'-UTR sequences suggests that they may play an important role in regulating gene expression. In this report, we showed that the 5'-UTR sequence of PTEN inhibits translation of its downstream



**Figure 10** Endogenous PTEN transcripts with shorter 5'-UTRs exist in human cancer cell lines. (a) Mapping 5'-UTRs of endogenous PTEN transcripts by 5' RACE. Total RNAs (2  $\mu$ g) from DU145 cell line were used for 5' RACE as described in Materials and methods. The final DNA products were separated by agarose gel electrophoresis (lane 2). The 5' end sequencing results of the two small products were shown on the right. (b) Mapping 5'-UTRs of endogenous PTEN transcripts by an RNase protection assay. The RNase protection assay with 10  $\mu$ g of RNAs from DU145 (lane 3) and H226 (lane 4) cells, 10  $\mu$ g of yeast tRNA (lane 2), and a probe covering the region of -483 and -60 in the 5' UTR sequence of human PTEN was performed as described in Materials and methods. The three distinctively protected products with estimated sizes of 360, 135, and 59 bases are indicated. Lane 1 shows the mobility and the estimated size of the probe used. (c) Schematic diagram showing relative locations of transcription start sites, the primer used in the 5' RACE, and the probe used for RPA. The asterisks show the possible corresponding positions in the PTEN sequence for the products generated from both 5' RACE (panel a) and RNase protection assay (panel b)

sequence in an *in vitro* system. A transcript with a 5'-UTR of only 93 bases is 100-fold more translatable than the transcript with a long 5'-UTR of 828 bases. The inhibitory effect of the long 5'-UTR is similar to that of the 5'-UTR of PDGF, which inhibits translation about 40-fold (Rao *et al.*, 1988).

Inhibition of translation by 5'-UTR led us to explore other mechanisms such as IRES that may be used for PTEN translation. The 5'-UTR sequence of PTEN enhanced the expression of the second cistron by about 70-fold in the traditional dicistronic DNA assays and translation of the second cistron persisted when a hairpin structure was used to block translation of the first cistron (Figure 3). However, no such enhancing activity was observed in the dicistronic RNA assay, indicating that the 5'-UTR of PTEN may not have IRES. This conclusion is supported by our promoterless dicistronic test, which revealed that the enhancing effect was due to the existence of promoter activities in the 5'-UTR of PTEN. In addition, Northern blot analysis clearly detected the monocistronic transcript derived from the promoter in the 5'-UTR sequence of PTEN

(Figure 7). By construction of a variety of deletion mutants from both 5' and 3' end of the 5'-UTR, we mapped the promoter activity to the region -551 to -220. Our experiments using *Drosophila* SL2 cells showed that the promoter in the 5'-UTR sequence of PTEN responds to Sp1, a ubiquitous transcription factor that is not only important for constitutive expression of many house-keeping genes but also critical for induced expression of many growth-related genes (Black *et al.*, 2001). Finally, using the RNase protection assay and 5' RACE, we identified two endogenous PTEN transcripts with short 5'-UTRs of 109 and 193 bases in both human cancer cell lines H226 and DU145, which were known to express human PTEN (Whang *et al.*, 1998; Soria *et al.*, 2002).

In the past, promoters in the 5'-UTR sequences have primarily been ignored and lack attention. Promoter studies normally begin with primer extension aiming at identification of the transcription start site furthest away from the translation initiation start codon. Subsequent analysis usually focuses on the promoter regions that are responsible for the generation of the transcript with the

longest 5'-UTR. Few of the past studies considered the impact of translational regulation and the fact that mRNAs with long 5'-UTRs generated by the promoter they studied may not be translatable. For the same reason, the reporter activity detected in the presence of both the upstream promoter and the long 5'-UTR sequence might be due to the contribution of another promoter in the 5'-UTR sequence that produces a transcript with a shorter 5'-UTR. Consequently, many promoters in 5'-UTRs were never discovered. Such a strategy may introduce confusion to the researchers in the translation field who are working to understand how mRNAs with long 5'-UTRs can be translated.

In the case of the PTEN promoter study, Stambolic *et al.* (2001), using a conventional approach, identified that the transcription start site for transcripts with the longest 5'-UTR was located between -951 and -925 bases upstream of the translation start codon. However, our studies showed that a transcript with more than 500 bases of 5'-UTR could not be translated efficiently and it does not appear to contain IRES. The question thus remains as to how the mRNA of PTEN with a 5'-UTR of 925-950 bases is translated. Use of the promoter in the 5'-UTR sequence to produce mRNAs with shorter 5'-UTRs would provide a solution to this problem. In fact, of the 100 EST clones found by searching the GeneBank using the 5'-UTR of mouse PTEN as a query sequence, 50 clones may have a 5'-UTR less than 500 bases. Our RACE and RPA study (Figure 10), promoter mapping (Figure 8), and Northern blot analysis (Figure 7) suggested the existence of transcripts with various lengths of 5'-UTR. Furthermore, using RPA and 5' RACE, we identified two endogenous transcripts with 5'-UTRs of 109 and 193 bases, respectively, that are likely originated from the 5'-UTR promoter. However, this possibility was not considered in the Stambolic *et al.* (2001) study and a large part (427 bp immediately upstream of the translation start codon) from the 5'-UTR sequence of PTEN was deleted in their constructs. By doing so, their promoter activity is only fourfold over the vector control as compared to about 100-fold shown in our study (Figure 6). They, therefore, failed to explain the p53-independent, relatively high and constant expression of PTEN mRNA and protein in a variety of cells.

In another study (Virolle *et al.*, 2001), the transcription start site at -1013 bases upstream of the translation start codon was considered simply based on the longest cDNA sequence of human PTEN. Nevertheless, the whole 5'-UTR sequence was used, which led to the identification of the Egr-1 factor binding site located downstream of the postulated transcription start site. However, the existence of promoters in the 5'-UTR sequence, which may direct alternative transcription compatible with translation, was not considered. Although not discussed in their study, their results are consistent with ours in the present study. Furthermore, they also showed that the promoter activity in the 5'-UTR is about half of that generated in the presence of both 5'-UTR and the upstream promoter. Again, these data were not discussed and they still considered in their

figures that the transcription start site is located at the upstream. This type of routine practice in promoter studies will likely generate incorrect conclusions on the regulation of gene expression, specifically for PTEN because it is known that the mRNAs of PTEN with long 5'-UTRs cannot be translated.

Our observations also confirm that the translation of PTEN mRNAs is regulated by the secondary structure of their 5'-UTRs. Only mRNAs with shorter 5'-UTRs that have less free energy in secondary structure can be significantly translated. Retrospectively, it is noteworthy that human PTEN transcripts of heterogeneous sizes between 2 and 5 kb are observed on Northern blot (Whang *et al.*, 1998). Our RPA and 5' RACE results (Figure 10) demonstrated that 5'-UTRs of endogenous PTEN mRNAs in human cancer cell lines are heterogeneous at least in their length with the 5'-UTRs. Interestingly, in the human PTEN promoter study (Virolle *et al.*, 2001), the  $\Delta 3'$  transcript with a short 5'-UTR (~258 bases) has a much higher basal and Egr-1-stimulated expression than the transcript with a longer 5'-UTR (~1031). Although these authors were focusing on the positive and negative elements responsible for transcriptional regulation of PTEN expression, it cannot be ruled out that the transcripts with a short 5'-UTR are in fact translated with higher efficiency than the transcripts with a long 5'-UTR. Furthermore, the relative abundance of PTEN mRNA does not correlate well with its protein level in advanced prostate cancer tissues as compared with normal tissues (Whang *et al.*, 1998), suggesting that translational control may play an important role in regulating PTEN expression.

Since the promoter in the 5'-UTR, which was mapped to the region between -551 and -220 bases upstream of the translation start codon, contributes to production of translation-compatible transcripts, and functions in a variety of cell lines tested (Figure 6), it is tempting to suggest that the promoter in the 5'-UTR is a major promoter that controls ubiquitous and constitutive PTEN protein production. Upstream promoters or enhancers may function through regulating the production of transcripts from the promoter in the 5'-UTR. Switching off the 5'-UTR promoter will, therefore, lead to reduced or complete loss of PTEN expression. Like many other tumor suppressor genes, expression of PTEN has been postulated to be regulated by epigenetic mechanisms. PTEN expression can be lost or greatly reduced in some tumors without any mutation in the coding sequence of the PTEN gene in endometrial, breast, prostate, ovarian, non-small cell lung cancer, and melanocytic tumors (reviewed in Mutter, 2001; Leslie and Downes, 2002; Soria *et al.*, 2002). This is supported by an observation that PTEN expression in the prostate cancer xenograft LuCap-35 and lung cancer cell line H1299 was recovered upon treatment with the demethylating agent 5-azadeoxycytidine (Whang *et al.*, 1998; Soria *et al.*, 2002). However, the detailed mechanism underlying epigenetic regulation of PTEN expression is yet to be identified. In this respect, it is interesting to note that the 5'-UTR-encoding sequence of both human and mouse PTEN contains two CpG islands, which may

be potential methylation targets as predicted using an online program (<http://www.ebi.ac.uk/emboss>) (data not shown). Indeed, the presence of methylation in the 5'-UTR sequences (-405 to -104) was reported to be frequent in both endometrial and gastric carcinomas (Salvesen *et al.*, 2001; Kang *et al.*, 2002). This region with high frequency of methylation overlaps with the region containing the promoter mapped in this study (-551 to -220). Methylation of this region may shut off 5'-UTR promoter activity and constitutive expression of PTEN and, therefore, inactivate the expression of PTEN in some tumors. We are currently testing this hypothesis by determining the effect of methylation of the 5'-UTR sequence on promoter activity.

Since the promoter in the 5'-UTR sequence of PTEN may serve as a major promoter controlling constitutive expression of PTEN, detailed promoter mapping may lead to discovery of *cis*-elements and transcription factors that govern constitutive PTEN expression. Analysis of putative transcription factor binding sites showed that the promoter in the 5'-UTR sequence contains several binding sites for transcription factors such as Sp1, Egr-1, Ets-1, Ap-1, Ap-2, and Ap-4. We have shown that Sp1 is involved in the activation of this promoter. Interestingly, the Sp1 site located at -82/-77 plays a critical role in the p53/p53 site (at 2281/2262)-mediated synergistic transactivation of the p21 promoter (Koutsodontis *et al.*, 2001). Similarly, conditional formation of transcriptional Sp1-p53 regulatory complexes has been reported recently in a number of other promoters (Borellini and Glazer, 1993; Gualberto and Baldwin, 1995; Ohlsson *et al.*, 1998; Torgeman *et al.*, 2001). Sp1/p53 synergism may serve as a general mechanism for transcriptional activation of p53 target genes. Therefore, it is tempting to hypothesize that the possible synergism between the p53 site located in -1190 to -1157 of human PTEN promoter and the multiple Sp1 sites located in the 5'-UTR region also function on PTEN promoter. On the other hand, multiple consensus Egr-1 sites (GCGGCGGCG) identified in human PTEN promoter (Virolle *et al.*, 2001) were also found in the 5'-UTR sequence of mouse PTEN (Figure 9a). Whether these Egr-1 sites are functional and whether the Sp1 and Egr-1 transcription factors interplay on the promoter in the 5'-UTR sequence of PTEN remain to be addressed in future studies.

## Materials and methods

### Materials

Restriction enzymes, m<sup>7</sup>GpppG cap analogue, and Pfu polymerase were purchased from New England Biolabs, Amersham/Pharmacia Biotech, and Stratagene, respectively. The SL2 cell line was from ATCC. Sp6 RNA polymerases, RNasin, RNase-free DNase, RRL, LUC assay 'Stop & Glo' kit, and pSP64 PolyA plasmid were from Promega. RNeasy Mini Kit and Oligotex mRNA Mini Kit were from Qiagen. Rediprime II Random Prime Labeling System,  $\alpha$ [<sup>32</sup>P]dCTP and  $\alpha$ [<sup>32</sup>P]CTP were from Amersham Biosciences. The Sephadex G-25 Quick Spin Columns (TE) for radiolabeled DNA and

RNA purification was from Roche Diagnostics. MAGNA nylon transfer membrane was from Osmonics Inc. Zero Blunt PCR Blunt PCR Cloning kit, Schneider's *Drosophila* culture medium, Lipofectamine plus and Lipofectin transfection reagent, and the 5'-RACE system for rapid amplification of cDNA ends were purchased from Invitrogen. Oligonucleotides were synthesized by Sigma-Genosys. IMAGE EST clones were obtained from either ATCC or Research Genetics. MAXI-script *in vitro* transcription kit and RPA III ribonuclease protection assay kit were products of Ambion. TaKaRa LA Taq polymerase with GC buffer was purchased from TaKaRa Bio Inc.

### Construction of plasmids

The cDNA clone mncb-0146 (GeneBank Accession #: AU035162) was from the Sugano mouse brain mncb library (Suzuki *et al.*, 2000). Double-strand DNA sequencing showed that this clone encodes the full-length mouse PTEN protein and confirmed that the 5'-UTR sequence (-828 to -1) was identical to that shown in GeneBank (Accession # NM\_008960). The cDNA, following sequencing, was cloned into pGEM-4Z (Promega) at *EcoRI* and *XbaI* sites, resulting in the plasmid pG-WPTEN (Figure 1a). In the deletion construct pG-DPTEN, the 5'-UTR sequence from -828 to -8 was removed by replacing the *EcoRI/BglII* fragment representing -828 to +318 with a PCR fragment representing -8 to +318. The plasmid pG-LUC was generated by cloning *EcoRI-XbaI* fragment of pRF (Stoneley *et al.*, 2000), corresponding to the firefly LUC-encoding region into pGEM-4Z. To generate pG-5'LUC, the -828 to -1 of the 5'-UTR sequence of mouse PTEN was amplified by PCR and inserted into pG-LUC between the *EcoRI* and *NcoI* sites. The 5'-UTR deletion mutants of mouse PTEN (-551/-1, -479/-1, -299/-1) were generated by cloning *XhoI-NcoI*, *NotI-NcoI*, *SmaI-NcoI* fragments into pG-LUC, respectively. The -257/-1 and -93/-1 deletion mutants were generated by PCR.

Dicistronic constructs were generated based on pRF. The 5'-UTR sequence of mouse PTEN (-828 to -1) from pG-LUC was cloned into pRF, resulting in pR-PTEN-F. In addition, a 492 bp of the 5'-UTR sequence of human PTEN cDNA (-492 to -1) was amplified from human IMAGE 2157760 (GeneBank Accession #: AI480306) and cloned into pRF, resulting in pR-hPTEN-F. Dicistronic constructs with a hairpin structure were generated by inserting into *EcoRV* site of pRF and pR-PTEN-F a double-stranded oligonucleotide with sense sequence of 5'-ATCAAAAGCGCAGGTCGCGACCGCGCATGCGCGGTCGCGACCTGCGCTAAAGAT-3'. The dicistronic constructs containing the truncated 5'-UTR of PTEN (-551/-1, -479/-1, -299/-1, -257/-1) were generated by cloning these cDNA fragments from the pG-LUC-based constructs (see above) into pRF. The dicistronic constructs containing other truncations (-828/-220, -828/-348, and -828/-474) were obtained by deletions from pR-PTEN-F between *BsrXI* and *NcoI*, *SmaI* and *NcoI*, and between *NotI* and *NcoI*. The -828/-40 construct was generated by cloning a PCR fragment into pRF. The promoterless dicistronic constructs were generated as previously described (Han and Zhang, 2002).

Constructs containing poly(A) for *in vitro* transcription were engineered using the vector pSP64 PolyA (Promega) that has 30 bp (dA-dT) sequence. The *EcoRV-XbaI* fragment of the pRF vector containing the *Renilla* LUC gene was first cloned into pSP64 PolyA vector at the *XbaI* and blunted *HindIII* sites to generate the plasmid pSP-R<sub>A30</sub>. The *XbaI* fragment of pR-HRV-F containing the IRES of HRV and the firefly LUC gene



was then isolated and cloned into pSP-R<sub>A30</sub> at the *Xba*I site to generate pSP-R-HRV-F<sub>A30</sub>. The pSP-RF<sub>A30</sub> plasmid was obtained by removing the IRES sequence of HRV from the pSP-R-HRV-F<sub>A30</sub> by digestion with *Spe*I and *Nco*I. To engineer pSP-R-PTEN-F<sub>A30</sub>, the *Xho*I-*Nco*I fragment containing -551 to -1 sequence from pR-PTEN-F was used to replace the HRV IRES fragment in pR-HRV-F<sub>A30</sub> construct. All the above constructs were confirmed by double-strand DNA sequencing.

#### *In vitro* transcription and translation

*In vitro* transcription and translation were performed as previously described (Zhang and Ling, 1991). The pG-WPTEN-, pG-DPTEN-, and pG-LUC-based plasmids were linearized by *Xba*I and pRF<sub>A30</sub>-based plasmids were linearized by *Eco*RI. The capped transcripts were synthesized in the presence of 1 mM m<sup>7</sup>GpppG and purified using a Qiagen RNeasy Mini kit. Capped RNA transcripts (5–50 ng) were used to program cell-free translation in RRL in a final volume of 10  $\mu$ l containing 6.5  $\mu$ l RRL. Translation product was either visualized by autoradiography or by measurement of LUC activities.

#### Cell culture, DNA and RNA transfection

Human HeLa and HEK293 cells were cultured using DMEM, while H1299, K562, DU145, and H226 cells were maintained in RPMI1640, supplemented with 10% fetal bovine serum at 37°C with 5% CO<sub>2</sub>. Schneider's *Drosophila* cell line 2 was maintained in Schneider's *Drosophila* medium supplemented with 10% fetal bovine serum at room temperature with atmospheric CO<sub>2</sub>.

DNA transfection of human cell lines was performed with Lipofectamine plus reagents according to the manufacturer's protocol. In a 24-well plate, approximately  $1 \times 10^5$  cells/well were plated and transfected with 0.4  $\mu$ g DNA. Cells were harvested 24 h following transfection for LUC assay. Transfection of *Drosophila* SL2 cells was performed according to the protocol as previously described (Han et al., 2001).

RNA transfection was performed using the cationic liposome-mediated method as previously described (Han and Zhang, 2002). Briefly, approximately  $2 \times 10^5$  cells/well were seeded onto six-well plates on the day before transfection. Cells were washed once with Opti-MEM I-reduced serum medium (GIBCO-BRL) and left in the incubator with some medium during preparation of the liposome-polynucleotide complexes. Opti-MEM I medium (1 ml) in a 12  $\times$  75 mm polystyrene snap-cap tube was mixed with 12.5  $\mu$ g of Lipofectin reagent and 5  $\mu$ g capped mRNA. The liposome-RNA-medium mixture was immediately added to cells. At 8 h following transfection, cells were harvested and processed for LUC analysis.

#### Northern blot analysis

Subconfluent H1299 cells in 10-cm plates were transfected with 4  $\mu$ g/plate constructs using Lipofectamine Plus. At 24 h following transfection, the total RNAs were extracted using an RNeasy Mini Kit. Residual plasmid DNA in the total RNA was digested with RNase-free DNase. The polyA mRNAs were then isolated from 250  $\mu$ g of total RNAs using an Oligotex mRNA Mini Kit. One-fifth of the mRNAs were separated in 1% agarose gels in the presence of formaldehyde and MOPS buffer and blotted onto MAGNA

nylon membranes. The blots were hybridized with a <sup>32</sup>P-labeled firefly LUC DNA probe (1656 bp), which was isolated by cleaving pRF with *Nco*I and *Xba*I and labeled using the Rediprime II Random Prime Labeling System.

#### 5' RACE and RNase protection assay

5' RACE was performed based on the protocol provided by the supplier. Briefly, 2  $\mu$ g total RNA from human prostate cancer cell line DU145 was used. The primer used for first-strand cDNA synthesis was GSP1A: 5'-TGCGGGTGCTA-TAATGT-3', located in 238 bases downstream of the translation start codon ATG. The primers used for first-round PCR are 5' RACE Abridged Anchor Primer from the supplier (36 bases), human PTEN-specific primer GSP2: 5'-CCATCCTCTTGATATCTCCTTTTG-3', located 58 bases downstream of the translation start codon ATG. The primers used for the second and final PCR are Abridged Universal Amplification Primer (AUAP) from the supplier and PTEN-specific primer PTENEXT1: 5'-CCTGTGGCTGAAGAAAA-AGGAGGAGAGAGAT-3', located at 10 bases upstream the translation start codon ATG. Takara *Taq* polymerase and GC buffer II are used for PCR amplification. The reaction conditions were 94°C for 2 min followed by 40 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min. PCR products were purified from gel, blunted with Klenow enzyme, and then cloned using the PCR Zero Blunt Cloning kit. Individual clones were sequenced for determination of the transcription start sites.

Ribonuclease Protection Assay (RPA) was performed using the RPAIII kit according to the supplier's instructions. Briefly, the RNA probe was produced by first cloning the region of -770 to -60 of human PTEN 5'-UTR sequence into PCR-Blunt vector at *Spe*I and *Xba*I sites. The resulting plasmid was linearized with *Not*I and transcribed using T7 RNA polymerase in the presence of 0.5 mM each of ATP, GTP, UTP, and 0.01 mM CTP supplemented with 3.12  $\mu$ M a[<sup>32</sup>P] CTP. The <sup>32</sup>P-labeled probe was digested with DNase and purified using a Sephadex G-25 Quick Spin Column. About  $0.5 \times 10^6$  c.p.m. of probe was hybridized to 10  $\mu$ g total RNA at 45°C overnight followed by digestion with RNase T1/A at 37°C. The reaction was then stopped and the protected RNAs precipitated before separation by electrophoresis on a 6% acrylamide/8 M urea gel for autoradiography.

#### Abbreviations

SL2, Schneider's *Drosophila* cell line 2; kb, kilobase(s); PCR, polymerase chain reaction; 5'-UTR, 5'-untranslated region; IRES, internal ribosome entry site; PAGE, polyacrylamide gel electrophoresis; mRNA, messenger RNA; RRL, rabbit reticulocyte lysate; HRV, human rhinovirus; LUC, luciferase; VEGF, vascular endothelial growth factor; PDGF, platelet-derived growth factor.

#### Acknowledgements

We are indebted to Dr AE Willis for pRF and pGL3RHRV plasmids, Drs Guntram Suske and Etty Benveniste for pPacO and pPacSp1 plasmids, and Dr Philip Marsden for pPac-UEts-1 plasmid. This work was supported in part by National Institutes of Health Grants CA64539 and GM59475, and by Department of Defense Grant DAMD17-02-1-0073. J-TZ is a recipient of a Career Investigator Award from the American Lung Association.

## References

- Ali IU, Schriml LM and Dean M. (1999). *J. Natl. Cancer Inst.*, **91**, 1922–1932.
- Backman SA, Stambolic V, Suzuki A, Haight J, Elia A, Pretorius J, Tsao MS, Shannon P, Bolon B, Ivy GO and Mak TW. (2001). *Nat. Genet.*, **29**, 396–403.
- Bernstein J, Sella O, Le SY and Elroy-Stein O. (1997). *J. Biol. Chem.*, **272**, 9356–9362.
- Black AR, Black JD and Azizkhan-Clifford J. (2001). *J. Cell Physiol.*, **188**, 143–160.
- Borellini F and Glazer RI. (1993). *J. Biol. Chem.*, **268**, 7923–7928.
- Cantley LC and Neel BG. (1999). *Proc. Natl. Acad. Sci. USA*, **96**, 4240–4245.
- Coldwell MJ, Mitchell SA, Stoneley M, MacFarlane M and Willis AE. (2000). *Oncogene*, **19**, 899–905.
- Courey AJ and Tjian R. (1988). *Cell*, **55**, 887–898.
- Dahia PL. (2000). *Endocr. Relat. Cancer*, **7**, 115–129.
- Dennig J, Beato M and Suske G. (1996). *EMBO. J.*, **15**, 5659–5667.
- Di Cristofano A and Pandolfi PP. (2000). *Cell*, **100**, 387–390.
- Gan W, LaCelle M and Rhoads RE. (1998). *J. Biol. Chem.*, **273**, 5006–5012.
- Gan W and Rhoads RE. (1996). *J. Biol. Chem.*, **271**, 623–626.
- Groszer M, Erickson R, Scripture-Adams DD, Lesche R, Trumpp A, Zack JA, Kornblum HI, Liu X and Wu H. (2001). *Science*, **294**, 2186–2189.
- Gualberto A and Baldwin Jr AS. (1995). *J. Biol. Chem.*, **270**, 19680–19683.
- Han B, Liu N, Yang X, Sun HB and Yang YC. (2001). *J. Biol. Chem.*, **276**, 7937–7942.
- Han B and Zhang JT. (2002). *Mol. Cell. Biol.*, **22**, 7372–7384.
- Hisatake J, O'Kelly J, Uskokovic MR, Tomoyasu S and Koeffler HP. (2001). *Blood*, **97**, 2427–2433.
- Huez I, Creancier L, Audigier S, Gensac MC, Prats AC and Prats H. (1998). *Mol. Cell. Biol.*, **18**, 6178–6190.
- Kang YH, Lee HS and Kim WH. (2002). *Lab. Invest.*, **82**, 285–291.
- Koutsodontis G, Tentes I, Papakosta P, Moustakas A and Kardassis D. (2001). *J. Biol. Chem.*, **276**, 29116–29125.
- Kozak M. (1987). *Nucleic Acids Res.*, **15**, 8125–8148.
- Kozak M. (1991). *J. Cell. Biol.*, **115**, 887–903.
- Kozak M. (2001). *Mol. Cell. Biol.*, **21**, 1899–1907.
- Kwon CH, Zhu X, Zhang J, Knoop LL, Tharp R, Smeyne RJ, Eberhart CG, Burger PC and Baker SJ. (2001). *Nat. Genet.*, **29**, 404–411.
- Leslie NR and Downes CP. (2002). *Cell Signal.*, **14**, 285–295.
- Li DM and Sun H. (1997). *Cancer Res.*, **57**, 2124–2129.
- Li G, Robinson GW, Lesche R, Martinez-Diaz H, Jiang Z, Rozenfurt N, Wagner KU, Wu DC, Lane TF, Liu X, Hennighausen L and Wu H. (2002). *Development*, **129**, 4159–4170.
- Maehama T and Dixon JE. (1999). *Trends Cell Biol.*, **9**, 125–128.
- Mathews DH, Sabina J, Zuker M and Turner DH. (1999). *J. Mol. Biol.*, **288**, 911–940.
- Mutter GL. (2001). *Am. J. Pathol.*, **158**, 1895–1898.
- Ohlsson C, Kley N, Werner H and LeRoith D. (1998). *Endocrinology*, **139**, 1101–1107.
- Pestova TV, Kolupaeva VG, Lomakin IB, Pilipenko EV, Shatsky IN, Agol VI and Hellen CU. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 7029–7036.
- Quandt K, Frech K, Karas H, Wingender E and Werner T. (1995). *Nucleic Acids Res.*, **23**, 4878–4884.
- Rao CD, Pech M, Robbins KC and Aaronson SA. (1988). *Mol. Cell. Biol.*, **8**, 284–292.
- Ross AH, Lachyankar MB and Recht LD. (2001). *Neuroscientist*, **7**, 278–281.
- Sachs AB. (2000). *Cell*, **101**, 243–245.
- Salvesen HB, MacDonald N, Ryan A, Jacobs IJ, Lynch ED, Akslen LA and Das S. (2001). *Int. J. Cancer*, **91**, 22–26.
- Schneider R and Kozak M. (2001). *Mol. Cell. Biol.*, **21**, 8238–8246.
- Soria JC, Lee HY, Lee JI, Wang L, Issa JP, Kemp BL, Liu DD, Kurie JM, Mao L and Khuri FR. (2002). *Clin. Cancer Res.*, **8**, 1178–1184.
- Stambolic V, MacPherson D, Sas D, Lin Y, Snow B, Jang Y, Benchimol S and Mak TW. (2001). *Mol. Cell*, **8**, 317–325.
- Stambolic V, Suzuki A, de la Pompa JL, Brothers GM, Mirtsos C, Sasaki T, Ruland J, Penninger JM, Siderovski DP and Mak TW. (1998). *Cell*, **95**, 29–39.
- Stoneley M, Subkhankulova T, Le Quesne JP, Coldwell MJ, Jopling CL, Belsham GJ and Willis AE. (2000). *Nucleic Acids Res.*, **28**, 687–694.
- Suzuki Y, Ishihara D, Sasaki M, Nakagawa H, Hata H, Tsunoda T, Watanabe M, Komatsu T, Ota T, Isogai T, Suyama A and Sugano S. (2000). *Genomics*, **64**, 286–297.
- Torgeman A, Mor-Vaknin N, Zelin E, Ben-Aroya Z, Lochelt M, Flugel RM and Aboud M. (2001). *Virology*, **281**, 10–20.
- Virolle T, Adamson ED, Baron V, Birle D, Mercola D, Mustelin T and de Belle I. (2001). *Nat. Cell Biol.*, **3**, 1124–1128.
- Whang YE, Wu X, Suzuki H, Reiter RE, Tran C, Vessella RL, Said JW, Isaacs WB and Sawyers CL. (1998). *Proc. Natl. Acad. Sci. USA*, **95**, 5246–5250.
- Willis AE. (1999). *Int. J. Biochem. Cell Biol.*, **31**, 73–86.
- Zhang JT and Ling V. (1991). *J. Biol. Chem.*, **266**, 18224–18232.

## Tight Control of Platelet-derived Growth Factor B/c-sis Expression by Interplay between the 5'-Untranslated Region Sequence and the Major Upstream Promoter\*

Received for publication, May 12, 2003, and in revised form, August 29, 2003  
Published, JBC Papers in Press, September 5, 2003, DOI 10.1074/jbc.M304976200

Baoguang Han, Zizheng Dong, and Jian-Ting Zhang†

From the Department of Pharmacology and Toxicology, Walther Oncology Center/Walther Cancer Institute and Indiana University Cancer Center, School of Medicine, Indiana University, Indianapolis, Indiana 46202

The long and GC-rich 5'-untranslated region (5'-UTR) of the known 3.8-kb platelet-derived growth factor B (PDGF-B)/c-sis mRNA is highly conserved and inhibits its own translation. It has been thought that this 5'-UTR functions by regulating translation possibly using an internal ribosome entry site (IRES)-mediated mechanism. However, in the present study we found no evidence that the 5'-UTR sequence of PDGF-B mRNA contains any IRES activity. Instead, we found that the 5'-UTR sequence of PDGF-B functions as a promoter both constitutively and upon induction in a variety of cell lines. The 5'-UTR sequence contains two promoters (termed P1 and P2) when only the 5'-UTR sequence is analyzed. In the presence of the upstream TATA-box-containing promoter (P0), P1 and P0 promoters are integrated into one promoter, whereas the P2 promoter still functions. The full promoter with combined P0, P1, and P2 produced two transcripts, with the major one having the full-length 5'-UTR and the minor one the short 5'-UTR. The integrated P0/P1 promoter and P2 promoter are likely responsible for producing the endogenous 3.8- and 2.8-kb PDGF-B mRNAs that are detected in cultured human renal microvascular endothelial cells, a few tumor cells, and rat brain tissues. Furthermore, we detected the 2.8-kb PDGF-B mRNA in erythroleukemia K562 cells upon 12-*O*-tetradecanoylphorbol-13-acetate-induced differentiation. Considering that the 5'-UTR in the 3.8-kb mRNA contains no IRES activity and inhibits cap-dependent translation, we believe that the endogenous 2.8-kb mRNA produced from the 5'-UTR promoter is likely the major template responsible for protein production both constitutively and upon induction. We also found that the transcription from the 5'-UTR P2 promoter might be coordinated by the major upstream P0 promoter upon stimulation. Based on these observations, we propose that the TATA-containing P0 promoter and the 5'-UTR promoter work together to tightly control the expression of PDGF-B.

Platelet-derived growth factor B (PDGF-B)<sup>1</sup>/c-sis belongs to a family of proteins consisting of four gene products, PDGF-A, PDGF-B, PDGF-C, and PDGF-D. They function as homodimers or heterodimers to selectively signal through cell surface tyrosine kinase receptors (PDGFR- $\alpha$  and PDGFR- $\beta$ ) and regulate diverse cellular functions (1–3). PDGF in general has been implicated to play an important role in embryogenesis, wound healing, as well as in the development of several serious disorders, including certain malignancies, atherosclerosis, and various fibrotic conditions. Understanding the mechanism of regulation of PDGF expression is, thus, important for designing strategies to control the expression of PDGF in such disorders.

The regulation of human PDGF-B expression is complex. The human PDGF-B gene contains seven exons spanning 24 kb of the genomic DNA on human chromosome 22. Transcription is normally driven by a short basal TATA-containing promoter that is responsible for production of the regular 3.8-kb transcript (4–7). However, this mRNA contains a GC-rich 5'-untranslated region (5'-UTR) of 1022 bases with three AUG codons and a highly stable secondary structure. This type of 5'-UTR sequences normally have an inhibitory effect on translation. Indeed, it has been found that the high level of the 3.8-kb PDGF-B mRNA is not accompanied by detectable proteins in many cell lines (8–10). In contrast, the PDGF-B protein is detected in some cell lines despite the presence of low levels of the 3.8-kb mRNAs (11). An alternative 2.8-kb transcript that lacks the long 5'-UTR sequence was detected in a few tumor cell lines and was shown to be associated with high level of PDGF-B protein (12). The 2.8-kb mRNA was also reported in cultured human renal microvascular endothelial cells (HRMECs) upon stimulation with transforming growth factor  $\beta$  (TGF- $\beta$ ) or phorbol 12-myristate 13-acetate (13). Interestingly, the 2.8-kb mRNA is degraded in a protein synthesis-dependent pathway and can be selectively enriched by cycloheximide treatment. The 2.8-kb mRNAs have also been detected in rat brain at certain stages of brain development, and its detection was associated with the increased level of PDGF-B protein (12). Thus, it has been suggested that the expression of PDGF-B mRNA with short 5'-UTR sequences possibly exists widely in non-transformed tissues *in vivo* (12).

One of the major sites of PDGF synthesis is within bone marrow megakaryocytes, the platelets progenitor cells. Be-

\* This work was supported in part by National Institutes of Health Grants CA64539 and GM59475, and by Department of Defense Grant DAMD17-02-1-0073. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† A recipient of a Career Investigator Award from the American Lung Association. To whom correspondence should be addressed: Dept. of Pharmacology and Toxicology, IUCC, Indiana University School of Medicine, 1044 W. Walnut St., R4-166, Indianapolis, IN 46202. Tel.: 317-278-4503; Fax: 317-274-8046; E-mail: jianzhan@iupui.edu.

<sup>1</sup> The abbreviations used are: PDGF, platelet-derived growth factor; SL2, Schneider's *Drosophila* cell line 2; 5'-UTR, 5'-untranslated region; IRES, internal ribosome entry site; HRV, human rhinovirus; LUC, luciferase; TPA, 12-*O*-tetradecanoylphorbol-13-acetate; HRMEC, human renal microvascular endothelial cell; TGF- $\beta$ , transforming growth factor  $\beta$ ; RACE, rapid amplification of cDNA ends; CMV, cytomegalovirus; Mops, 4-morpholinepropanesulfonic acid; RPA, ribonuclease protection assay; CRD, coding region instability determinant.

cause human erythroleukemia K562 cells are differentiated into megakaryocytes upon TPA stimulation, regulation of PDGF-B expression in K562 cells has been extensively studied. The 3.8-kb species of PDGF-B mRNA was dramatically increased upon TPA-induced differentiation (14). Cis-elements and transcription factors, which are responsible for such induction, were also identified for the TATA-containing promoter (15, 16). Recently, the 5'-UTR of the 3.8-kb transcript has been reported to contain an internal ribosome entry site (IRES), which became more active in K562 cells upon differentiation into megakaryotic cells (17, 18). The IRES-mediated translation initiation works by directly recruiting the translational machinery to the nearby AUG translation start codon and therefore effectively avoids the inhibition of the long structured 5'-UTRs (reviewed in Refs. 19–21). However, whether this IRES really exists is questionable due to the inevitable technical problem associated with the traditional dicistronic test (22, 23). Like most of the previous cellular IRES studies, the 5'-UTR of PDGF-B is claimed to contain IRES based on transfection of artificial dicistronic plasmid followed by Northern blot analysis. Due to its low sensitivity, the use of Northern blot cannot differentiate IRES from cryptic promoter or potential splicing activities, all of which generate the same effect in dicistronic DNA test (25).

Herein, we report our surprising finding that the 2.8-kb human PDGF-B mRNA, previously detected in HRMECs (13), is also produced in K562 cells upon TPA-induced differentiation. This transcript represents a minor species in Northern blot analysis. However, treatment of cycloheximide stabilizes the 2.8-kb transcript and enables its clear detection by Northern blot, 5' rapid amplification of cDNA ends, and RNase protection assay. Using luciferase reporter promoter assay and Northern blot analysis, we demonstrated that the DNA sequence encoding the 5'-UTR of the long PDGF-B mRNA contains promoters that function in various cell lines and produces two mRNA species with a medium and a short 5'-UTR, respectively. However, in the presence of the upstream TATA-containing promoter, only the transcript with a short 5'-UTR was produced in addition to the transcript with the full-length 5'-UTR, which may explain why only the 3.8-kb and the 2.8-kb endogenous mRNAs are produced. On the other hand, the 5'-UTR sequence of the long PDGF-B transcript severely inhibits translation and does not display any IRES activity when tested using more stringent methods such as RNA transfection and promoterless dicistronic assay. Therefore, the 3.8-kb mRNA may contribute little, if any, to the production of PDGF-B protein. Based on the above observations, we conclude that the major TATA-containing promoter and the promoters in the 5'-UTR work together to control the expression of the 2.8-kb PDGF-B transcript in a variety of cells as an effective source of mRNA for protein production. Tight control of the production of the 2.8-kb mRNA and its stability may be used widely to control PDGF-B expression, both constitutively and upon stimulation.

#### EXPERIMENTAL PROCEDURES

**Materials**—Restriction enzymes were purchased from New England Biolabs. Site-directed mutagenesis QuikChange XL kit and Pfu polymerase were from Stratagene. Sp6 RNA polymerases, RNasin, RNase-free DNase, rabbit reticulocyte lysate, Luciferase assay 'Stop & Glo' kit, and pSP64 Poly(A) plasmid were from Promega. RNeasy Mini Kit and Oligotex mRNA Mini Kit were from Qiagen. The m<sup>7</sup>GpppG cap analogue, Rediprime II Random Prime Labeling System, [<sup>32</sup>P]dCTP and [<sup>32</sup>P]CTP were from Amersham Biosciences. The Sephadex G-25 Quick Spin Columns for purification of labeled DNA were from Roche Applied Science. MAGNA nylon transfer membrane was from Osmonics Inc. The 5'-RACE system for rapid amplification of cDNA ends and Lipofectin transfection reagents were purchased from Invitrogen. Oligonucleotides were synthesized by Sigma-Genosys. TaKaRa LA Taq polym-

erase was purchased from Takara Bio Inc. The Galacto-Light Plus chemiluminescent reporter assay kit for  $\beta$ -galactosidase was purchased from Tropix. The pTR-GAPDH-Human (pTR-glyceraldehyde-3-phosphate dehydrogenase-Human), the RNA Century Marker, the MAXI-script *in vitro* transcription kit, and RPA III ribonuclease protection assay kit were products of Ambion. Cycloheximide and TPA were from Sigma. The plasmid pSM1 (24) containing PDGF-B DNA insert was obtained from ATCC. The human IMAGE cDNA (ID 5174750 and GenBank™ accession number BC029822) containing the partial 5'-UTR and the entire coding region of PDGF-B were purchased from Open Biosystems.

**Construction of Plasmids**—The plasmid pBLHHCAT was a gift from Dr. William E. Fahl (McArdle Laboratory for Cancer Research, University of Wisconsin Medical School, Madison, WI). The plasmid contains a 5.5-kb PDGF-B DNA fragment (HindIII-HindIII), which includes the full-length 5'-UTR and the proximal promoter (5). The DNA sequence for the long 5'-UTR was amplified using the following primers: JB7, 5'-CCCCACTAGTGGCACTTCTCTCC and OS35, 5'-CCCCCATG-GCGACTCCGGGCGGCGCC (18). The purified PCR product was cloned into pRF, and pRF(-P) vector (25), resulting in pR-PDGF-F and pR-PDGF-F(-P), respectively. To construct the plasmid containing the 5' region, -807 to +1022, of PDGF-B (numbered relative to the known transcription start site in the remaining text unless otherwise specified), pHHCAT was first digested with HpaI and XhoI to isolate a 1.2-kb fragment that includes -807 to +475, which was then used to replace the SpeI/XhoI fragment in pR-PDGF-F(-P). To obtain the pRF(-P)-based construct that contains -807 to +82 of PDGF-B 5' region, a NheI/AvrII fragment isolated from the -807 to +1022 construct was cloned into pRF(-P) at the NheI/NcoI sites.

Mutagenesis of the TATA box was performed using the Stratagene QuikChange XL mutagenesis kit according to the manufacturer's protocol. The sense primer for mutagenesis was 5'-CGCACTCTCCCT-TCTCCCTCTAGATGGCCGGAACAGCTGAAAG-3' (with mutations shown underlined and in boldface). This primer changed the palindrome TTTATAAA sequence (the overlapping TATA box in both sense and antisense strand) into CTCTAGAT sequence that bears an XbaI site but shows no known binding sites for any transcription factors.

The 5'-end deletions of the 5'-UTR were generated using PCR for constructs containing the sequence from +75, +150, +225, +300, +375, +525, +600, or +675, to +1022. The other constructs containing sequence from +475, +769, or +862, to +1022 were obtained by deleting the SpeI/XhoI, SpeI/BamHI, and SpeI/SmaI fragments from the pR-PDGF-F(-P), respectively. The 3'-end deletion constructs containing +1 to +300, +1 to +395, +1 to +475, and +1 to +769 were obtained by deleting the RsrII/NcoI, SmaI/NcoI, XhoI/NcoI, and BamHI/NcoI fragments from pR-PDGF-F(-P), respectively. The construct containing +1 to +675 region of the 5'-UTR was obtained using PCR.

Dicistronic constructs containing poly(A) for *in vitro* transcription (Fig. 4) were engineered using the vector pSP64 Poly(A), which has 30-bp dA:dT. The EcoRV-XbaI fragment of the pRF vector containing *Renilla* luciferase gene was first cloned into pSP64 Poly(A) vector at the XbaI and blunted HindIII sites to generate the plasmid pSP-R<sub>A30</sub>. The XbaI fragment of pR-HRV-F (25) containing the IRES of HRV and the firefly luciferase gene was then isolated and cloned into pSP-R<sub>A30</sub> at the XbaI site to generate pSP-R-HRV-F<sub>A30</sub>. The pSP-RF<sub>A30</sub> plasmid was obtained by removing the IRES sequence of HRV from the pSP-R-HRV-F<sub>A30</sub> by digestion with SpeI and NcoI. To engineer pSP-R-PDGF-F<sub>A30</sub>, the SpeI-NcoI fragment from pR-PDGF-F, which contains the +1 to +1022 sequence, was used to replace the HRV IRES fragment in pR-HRV-F<sub>A30</sub> construct.

Monocistronic constructs containing poly(A) were also obtained from pSP64 poly(A). First, the XbaI fragment containing the 5'-UTR sequence of PDGF-B and the firefly luciferase-encoding sequence from pR-PDGF-F was cloned into pSP64 Poly(A), resulting in pSP-PDGF-F<sub>A30</sub>. Deletion of SpeI/BamHI, SpeI/SmaI, and SpeI/MluI regions from pSP-PDGF-F<sub>A30</sub> resulted in pSP-R-769-F<sub>A30</sub>, pSP-R-862-F<sub>A30</sub>, and pSP-R-983-F<sub>A30</sub>, which contain the +769 to +1022, +862 to +1022, and +983 to +1022 sequences of the 5'-UTR of PDGF-B, respectively. The pSP-R-375-F<sub>A30</sub> construct, which contains the +375 to +1022 region of the 5'-UTR, was constructed by cloning a PCR fragment into the SpeI/NcoI site of pSP-PDGF-F<sub>A30</sub>. All the above constructs were confirmed by DNA sequencing.

**In Vitro Transcription and Translation**—*In vitro* transcription and translation were performed as previously described (25). The pSP64 poly(A)-based dicistronic and monocistronic plasmids were linearized by using EcoRI. The capped transcripts were synthesized using the Sp6 RNA polymerase transcription system in the presence of 1 mM m<sup>7</sup>GpppG. DNA templates were digested using DNase, and the *in vitro*

transcripts were purified using a Qiagen RNeasy Mini kit and quantified. For *in vitro* translation, 25 ng of the capped RNA transcripts was used to program translation in rabbit reticulocyte lysate in a final volume of 10  $\mu$ l. Translation products were measured for firefly luciferase activities.

**Cell Culture, DNA, and RNA Transfection**—HeLa cells were maintained in Dulbecco's modified Eagle's medium, whereas H1299 and K562 cells were maintained in RPMI 1640 media both supplemented with 10% fetal bovine serum at 37 °C with 5% CO<sub>2</sub>.

DNA transfection in both HeLa and H1299 cells were performed with LipofectAMINE plus reagents according to the manufacturer's protocol. In a 24-well plate,  $\sim 1 \times 10^5$  cells/well were plated and transfected with 0.4  $\mu$ g of DNA. Cells were harvested 24 h following transfection for luciferase assay.

RNA transfection was performed using the cationic liposome-mediated method as previously described (25). Briefly,  $\sim 2 \times 10^5$  cells/well were seeded onto 6-well plates on the day before transfection. Cells were washed once with Opti-MEM I-reduced serum medium (Invitrogen) and left in the incubator with some medium during preparation of the liposome-poly nucleotide complexes. One milliliter of Opti-MEM I medium in a 12  $\times$  75-mm polystyrene snap-cap tube was mixed with 12.5  $\mu$ g of Lipofectin reagent and 5  $\mu$ g of capped mRNA. The liposome/RNA/medium mixture was immediately added to cells. Eight hours following transfection, cells were harvested and processed for luciferase analysis.

**Transient Transfection of K562 Cells**—Transient transfections of K562 cells were performed using electroporation. K562 cells were collected by centrifugation, washed with phosphate-buffered saline, and resuspended in the same buffer at  $10^7$  cells per 0.4 ml. For each electroporation, 0.4 ml of cell suspension was mixed with 20  $\mu$ g of constructs and 1.0  $\mu$ g of pCMV- $\beta$ gal in a final volume of 0.5 ml. Each electroporation pool received an electric pulse of 240 V and 1025 microfarads. After electroporation, the cells were incubated in 20 ml of prewarmed medium containing 20% serum for 24 h. The cells were then divided into two dishes with one supplemented with TPA (2 ng/ml) and other the same volume of solvent (ethanol) as a control. Forty-eight hours after TPA addition, the cells were collected by centrifugation, washed with phosphate-buffered saline, and lysed in passive lysis buffer and dual-luciferase, and  $\beta$ -galactosidase activity were determined.

**Northern Blot Analysis**—Subconfluent H1299 cells in 10-cm plates were transfected with 4  $\mu$ g/plate constructs using LipofectAMINE Plus. Twenty-four hours following transfection, the total RNAs were extracted using an RNeasy Mini kit and digested with RNase-free DNase to remove residual plasmid DNA. The poly(A) RNAs were then isolated from 250  $\mu$ g of total RNAs using an Oligotex mRNA Mini kit. One-fifth of the mRNAs were separated in 1% agarose gels in the presence of formaldehyde and MOPS buffer and blotted onto MAGNA nylon membranes. The blots were then hybridized with a <sup>32</sup>P-labeled firefly luciferase DNA probe (1656 bp), which was isolated by cleaving pRF with NcoI and XbaI and labeled using the Rediprime II random prime labeling system. The glyceraldehyde-3-phosphate dehydrogenase mRNA was also detected using specific probes as a control.

For Northern blot analysis of PDGF-B expression in K562 cells, the total RNA was prepared from normal and differentiated K562 cells. The differentiation of K562 cells into megakaryocytes was induced by TPA (2 ng/ml) treatment for 2 days. Cycloheximide (10  $\mu$ g/ml) was added to the medium on the third day following TPA treatment to stabilize the 2.8-kb mRNA. The poly(A) mRNA was isolated, and  $\sim 2$   $\mu$ g of mRNA of each sample was used for Northern blot analysis as described above. The probe used for Northern blot was a 2-kb fragment isolated from pSM1 (24) by BamHI digestion.

**PCR Analysis**—Determination of the 5'-end of the 2.8-kb species was performed with the 5' RACE system for rapid amplification of cDNA ends using the method provided by the supplier. Briefly, the first cDNA strand was synthesized using 5  $\mu$ g of total RNA isolated from K562 cells that was cultured with TPA for 2 days and subsequently cultured with both TPA and cycloheximide for another 16 h. The primer used for cDNA synthesis is human PDGF-B-specific antisense primer, 5'-CCACTGTCTCACACTTG-3', located at 548 bases downstream of the translation start codon. The primers used for first-round PCR were 5' RACE Abridged Anchor Primer from the supplier (36 bases) and human PDGF-B-specific antisense primer with a sequence of 5'-GCTTCTTC-CGCACAAT-3', located 488 bases downstream of the translation start codon. The primers used for the final round PCR were Abridged Universal Amplification Primer from the supplier and PDGF-B-specific primer, 5'-CTGACCAGACGCAGGTA-3', located at 56 bases downstream of the translation start codon. Takara Taq polymerase and GC

buffer II were used for PCR amplification. The reaction conditions were 94 °C for 2 min followed by 40 cycles of 94 °C for 1 min, 50 °C for 1 min, and 72 °C for 1 min. The PCR products were then purified, blunted with Klenow, and then cloned using the PCR Zero Blunt Cloning kit. Individual clones were isolated and sequenced to determine the transcription start sites.

Determination of the 5'-end of the longer transcript from the pR-PDGF-F(-P) transcription was performed by RT-PCR. About 5  $\mu$ g of total RNA isolated from pR-PDGF-F and pR-PDGF-F(-P)-transfected H1299 cells was used. The first strand of cDNA was synthesized using an antisense primer targeting firefly luciferase mRNA located 78 bases downstream of the translation start codon. PCR amplification was performed using antisense primer located at +675 and three sense primers located at +150, +225, +300, and +375 of the 5'-UTR of PDGF-B, respectively.

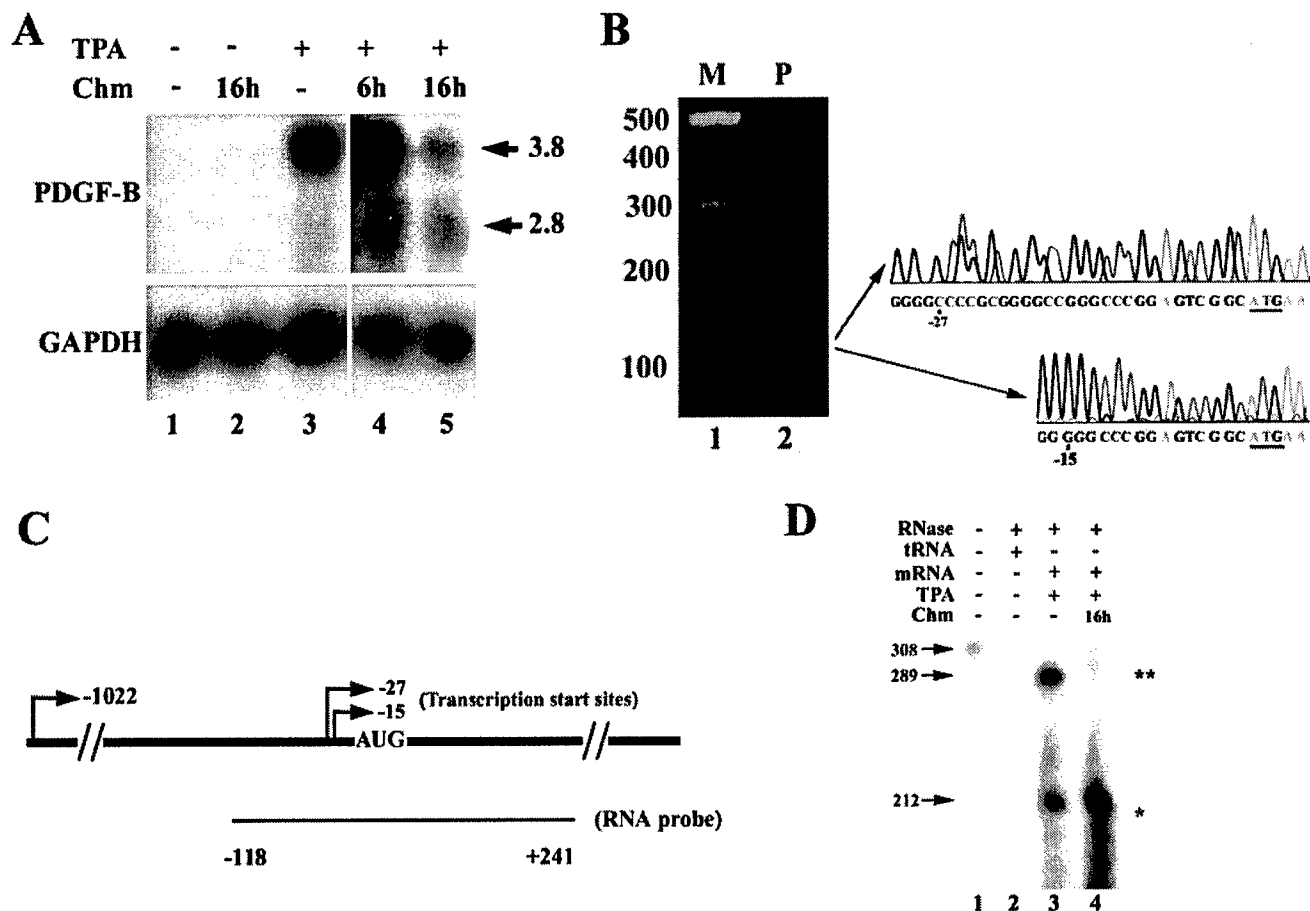
**Ribonuclease Protection Assay**—RPA was performed using the RPAIII kit according to the supplier's instructions. Briefly, the RNA probe was produced by first cloning into pGEM-4Z at the EcoRI and HindIII site the region -118 to +241 (numbered relative to the translation start site) of PDGF-B derived from the Image 5174750 (National Institutes of Health) cDNA clone by PCR. The resulting plasmid was linearized with EcoRI and transcribed using T7 RNA polymerase in the presence of 0.5 mM each of ATP, GTP, UTP, and 0.01 mM CTP supplemented with 3.12  $\mu$ M [ $\alpha$ -<sup>32</sup>P]CTP. The <sup>32</sup>P-labeled probe was digested with DNase and purified using a Sephadex G-25 Quick Spin column. About  $0.5 \times 10^6$  cpm of the probe was hybridized to 2  $\mu$ g of mRNA at 45 °C overnight followed by digestion with RNase T1/A at 37 °C. The reaction was then stopped, and the protected RNAs were precipitated before separation by electrophoresis on a 6% acrylamide/8 M urea gel for autoradiography.

## RESULTS

**Production of the 2.8-kb mRNA of PDGF-B in K562 Cells upon TPA-induced Differentiation**—Fen and Daniel (13) first described that the 2.8-kb PDGF-B mRNA with truncation at the 5'-UTR (it has a 5'-UTR with only 15 bases) was produced in HRMECs upon either TGF- $\beta$  or phorbol 12-myristate 13-acetate stimulation. It was thought that the 2.8-kb mRNA was degraded through a protein synthesis-dependent pathway, because treatment with cycloheximide selectively stabilized the 2.8-kb mRNA but not the 3.8-kb mRNA. Fen and Daniel also demonstrated that the 2.8-kb mRNA did not arise from post-transcriptional processing of the 3.8-kb mRNA but rather possibly from internal promoter within the 5'-UTR. The cycloheximide treatment did not enhance transcription of PDGF-B but rather increase the half-life of the 2.8-kb mRNA. The 5'-truncated PDGF-B mRNA was also observed in a few tumor cell lines and increased in rat brain at a certain stage of brain development, and the level of the 5'-truncated mRNA was shown to be associated with PDGF-B protein level (12).

We hypothesize that the 2.8-kb mRNA of PDGF-B may also be produced in K562 cells upon TPA-induced differentiation. To test this hypothesis, we first performed Northern blot analysis of the endogenous PDGF-B transcript in K562 cells and demonstrated that a major transcript of 3.8 kb was detected when K562 cells were treated with TPA for 2–3 days (Fig. 1A, lane 3). A minor transcript of 2.8 kb was also evident (lane 3) and more pronounced with longer exposure (data not shown). When cells were treated with cycloheximide for 6 or 16 h following TPA treatment, the 2.8-kb mRNA increased significantly (Fig. 1A, lanes 4 and 5). Treatment of K562 cells with cycloheximide alone, however, did not enhance expression of either mRNA species (Fig. 1A, lane 2), suggesting that cycloheximide alone did not affect PDGF-B transcription. Therefore, the TPA-stimulated K562 cells displayed a similar profile of PDGF-B transcription as HRMECs.

To map the 5'-end of the 2.8-kb PDGF-B mRNAs induced in differentiated K562 cells, we performed 5' RACE using RNAs isolated from K562 cells following TPA and cycloheximide treatment. After two rounds of PCR using nested PDGF-B-specific primers, we obtained a single PCR product (Fig. 1B,



**FIG. 1. Detection of the 2.8-kb PDGF-B mRNA in K562 cells upon TPA-induced differentiation.** **A**, Northern blot analysis to detect the 2.8-kb mRNA. K562 cells were treated with 2 ng/ml TPA for 2 days and then harvested (lane 3) or further co-treated with cycloheximide (10  $\mu$ g/ml) for additional 6 (lane 4) or 16 h (lane 5). The control K562 cells were either untreated (lane 1) or treated (lane 2) with cycloheximide alone for 16 h. Poly(A) RNAs were isolated for Northern blot analysis as described under "Experimental Procedures" using human PDGF-B cDNA probe. The 3.8- and 2.8-kb mRNAs are indicated by arrows. GAPDH probe was used as a control. **B**, 5'-RACE to map the transcription start site of the 2.8-kb mRNA. Total RNAs from K562 cells treated with TPA for 2 days followed by cotreatment with cycloheximide (10  $\mu$ g/ml) for 16 h were analyzed by 5'-RACE as described under "Experimental Procedures." The final DNA products were separated by agarose gel electrophoresis (lane 2) and cloned for sequencing (shown on the right). The translation start codon is underlined, and the transcription start sites are shown by the numbers. **C**, schematic diagram illustrating the location of transcription start sites and the probe used for RPA analysis. **D**, RPA analysis to determine the 5'-end of the 2.8-kb mRNA. The RNA probe shown in **C** was incubated with 10  $\mu$ g of yeast tRNA (lane 2), 2  $\mu$ g of mRNA from cells treated with TPA for 2 days (lane 3) and cells treated with TPA for 2 days followed by cotreatment with cycloheximide for 16 h (lane 4). The integrity of probe is shown in lane 1. The asterisks indicate the protected fragments. The size of the probe and protected fragments were estimated by linear regression using the RNA Century Marker from Ambion. Chm, cycloheximide.

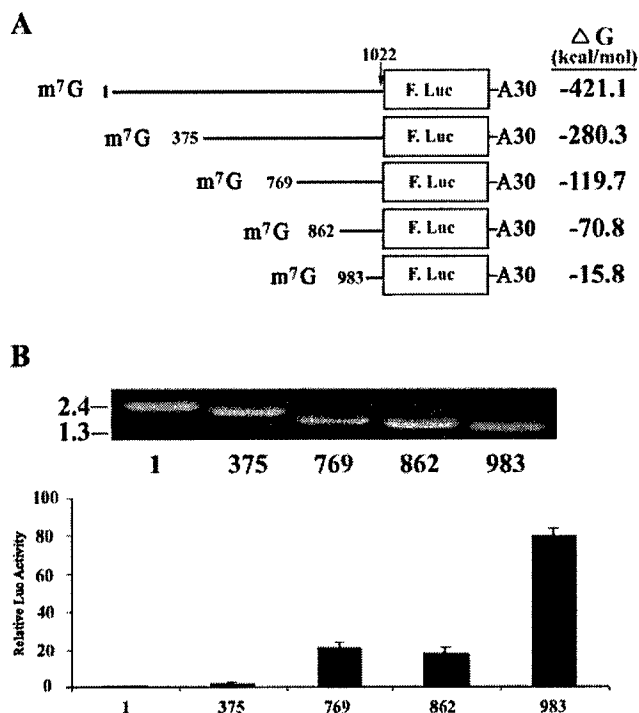
lane 2). The PCR product was cloned, and 10 individual clones were selected for sequencing. Three of the clones were identical and start at 15 bases upstream of the translation start codon ATG, whereas the other seven start at 27 bases upstream of the translation start codon (Fig. 1B). This observation suggests that there are at least two transcription start sites for the short RNA species of PDGF-B in differentiated K562 cells.

Because the results in 5'-RACE analysis could also be due to premature termination of polymerase reaction by the secondary structures in the 5'-UTR sequence, we performed a ribonuclease protection assay (RPA) using a PDGF-B probe spanning -118 and +241 (numbered relative to the translation start codon) (Fig. 1C). As shown in Fig. 1D, two fragments were found to be protected, and their sizes were estimated to be 289 and 212 bases, respectively (lane 3). While the longer protected fragment is possibly derived from the 3.8-kb transcript, the smaller fragment with ~77 bases less is likely derived from the 2.8-kb transcript with shorter 5'-UTRs. It is noteworthy that the smaller fragment represents a significant species of protected fragments in samples treated either without (lane 3) or with cycloheximide (lane 4), suggesting the 2.8-kb transcript

appears to be more efficiently protected. It also appears that the 2.8-kb transcript (the smaller fragment) is relatively more abundant upon cycloheximide treatment (compare lanes 3 and 4). Thus, these findings support the conclusion derived from the results shown in Fig. 1 (A and B) and confirm that the cycloheximide treatment selectively stabilizes the 2.8-kb mRNA.

**The Long 5'-UTR Sequence of PDGF-B Inhibits Cap-dependent Translation**—Although the 2.8-kb mRNA represents a minor species in TPA-induced PDGF-B transcripts, it may serve as the primary template for PDGF-B protein production, because the full-length PDGF-B transcript may not be efficiently translated due to translation inhibition by the long 5'-UTR. To test this concept, we determined the effect of the long 5'-UTR sequence of PDGF-B on translation. For this purpose, we constructed a series of plasmids that were used for generating *in vitro* transcripts with different lengths of 5'-UTRs upstream of a firefly luciferase reporter gene. These deletion mutants contain putative secondary structures with different free energies ( $\Delta G$ ) ranging from -421 to -15 kcal/mol as predicted by Zuker's mfold program (26) (Fig. 2A). The  $\Delta G$  value is an





**FIG. 2. Effect of deletions of the 5'-UTR sequence of PDGF-B on the translation of a heterologous reporter firefly luciferase.** A, schematic diagram of sequential deletions in the 5'-UTR sequence of human PDGF-B. The free energy for the predicted secondary structure of the 5'-UTR sequence is shown on the right (computed using MFOLD software (35)). The positions of the 5'-end of each deletion are indicated on the left. These 5'-UTRs and the firefly luciferase encoding regions were engineered into pSP64 Poly(A) for production of *in vitro* transcripts. B, *in vitro* transcription and translation. The deletion mutants were used to generate *in vitro* transcripts (upper panel), which were then used to program cell-free translation in rabbit reticulocyte lysate. Luciferase activities of the translated products were measured by enzymatic assays as described under "Experimental Procedures" and was normalized against that of the wild type 5'-UTR (bottom panel).

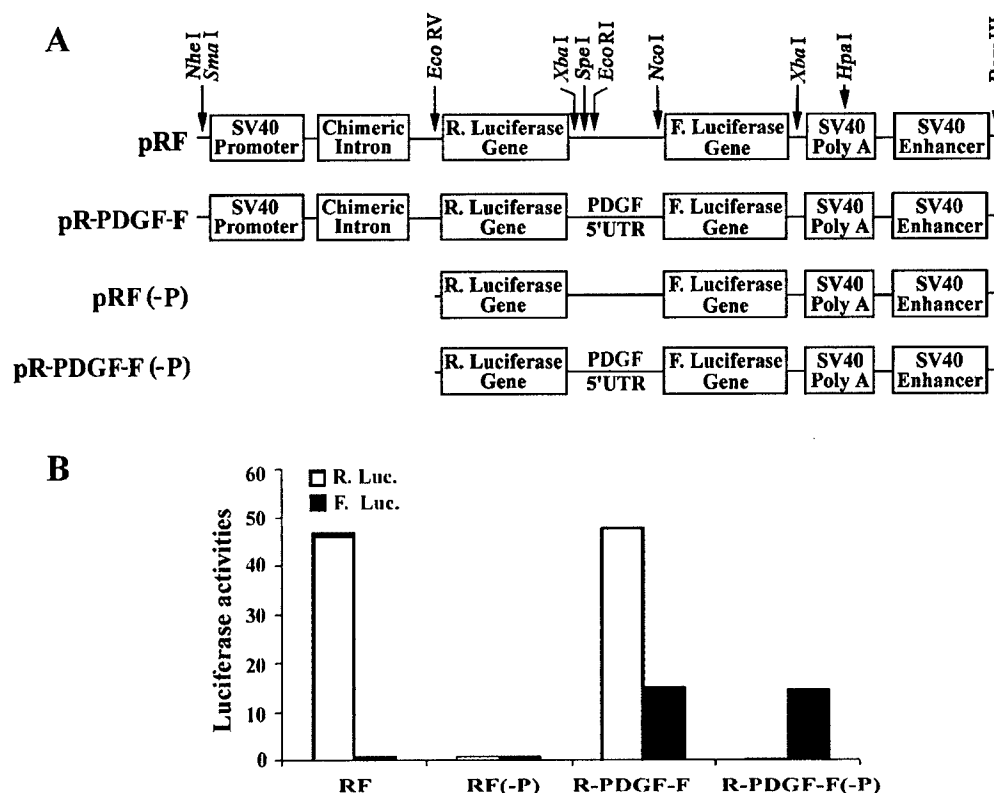
indicator of the stability of the secondary structure that may inhibit a ribosome-scanning process for cap-dependent translation (27). The higher the  $\Delta G$  value (with a minus sign), the more stable is the secondary structure. Capped *in vitro* transcripts were then generated (Fig. 2B), and equal amounts of the transcripts were used to program cell-free translation in rabbit reticulocyte lysate. The products were measured as firefly luciferase activity. As shown in Fig. 2B, the presence of the 5'-UTR with full-length (+1 to +1022) or 647 bases (+375 to +1022) severely inhibited the luciferase production. The transcript with a 5'-UTR of 253 bases (+769 to +1022) or 160 bases (+862 to +1022) is about 20-fold more efficiently translated than the one with a full-length 5'-UTR. Deletion of most of the 5'-UTR sequence (+983 to +1022) increased the translation by about 80-fold. These results suggest that the long 5'-UTR sequence of PDGF-B in the 3.8-kb transcript inhibits the translation of its downstream sequence, and the transcripts with short 5'-UTRs are much better templates for translation of a heterologous gene, consistent with previous findings (4, 28).

**The Long 5'-UTR Sequence Functions as a Promoter but Contains No Detectable Internal Ribosome Entry Site Activity**—It has been demonstrated that the 2.8-kb mRNA was not a product of post-transcriptional processing of the 3.8-kb mRNA but rather from independent transcription (12, 13). To determine whether the 5'-UTR sequence contains promoter activity, we cloned the 5'-UTR of PDGF-B into both pRF and pRF(-P) vector (25) (Fig. 3A). The pRF-based constructs contain an SV40 promoter to drive the transcription of dicistronic RNA

encoding *Renilla* luciferase as the first cistron and firefly luciferase as the second cistron. Translation of the first cistron serves as an indicator of cap-dependent translation, whereas translation of the second cistron reflects the IRES activity associated with the inserted intergenic sequence if a promoter or an alternative splicing site does not exist. The pRF(-P)-based constructs are derivatives of pRF by removing SV40 promoter and are useful for detecting promoter activity (Fig. 3A). Following transfection of these constructs into H1299 or HeLa cells, lysates were prepared for determination of both firefly (second cistron) and *Renilla* (first cistron) luciferase activities. As shown in Fig. 3B, the firefly luciferase activity was very low in H1299 cells transfected with pRF vector control compared with the *Renilla* luciferase, whereas insertion of the 5'-UTR sequence of PDGF-B into the intergenic region (pR-PDGF-F) enhanced the expression of firefly luciferase by ~17-fold as compared with the vector control. In cells transfected with the promoterless vector pRF(-P), both *Renilla* and firefly luciferase activities were low. However, pR-PDGF-F(-P)-transfected H1299 cells displayed similar firefly luciferase activity as pR-PDGF-F transfected H1299 cells, although there was no increase in *Renilla* luciferase activity. Similar results were also observed in experiments performed with HeLa cells (data not shown). These results strongly suggest that the long 5'-UTR sequence of PDGF-B contains promoter activity in both HeLa and H1299 cells.

Previously, it has been suggested that the long 5'-UTR sequence of PDGF-B contains an internal ribosome entry site (IRES), which can mediate cap-independent translation (17, 18). However, the data shown in Fig. 3 strongly suggest that the previously claimed IRES activity of PDGF-B is likely due to the promoter activity present in the 5'-UTR sequence. To further determine whether the long 5'-UTR sequence of PDGF-B contains IRES activity, dicistronic mRNAs with m<sup>7</sup>GpppG cap and polyadenylated tail were generated from dicistronic constructs (pR-PDGF-F<sub>A30</sub>) (Fig. 4A). Equal amounts of *in vitro* transcripts were then transfected into H1299 cells by Lipofectin encapsulation. Transfections using dicistronic RNAs from pRF<sub>A30</sub> vector and HRV-IRES-containing plasmid pR-HRV-F<sub>A30</sub> were performed as negative and positive controls, respectively. Eight hours following transfection, cell lysates were prepared for luciferase activity measurement. The efficiency of transfection was similar for all three RNAs as suggested by *Renilla* luciferase activity (with arbitrary units of 3360, 1183, and 1876, respectively, for RF<sub>A30</sub>, R-PDGF-F<sub>A30</sub>, and R-HRV-F<sub>A30</sub> in one typical experiment). As expected, RF<sub>A30</sub> displayed a very low level of firefly luciferase activity, which only represents about 0.06% of *Renilla* luciferase activity. The firefly luciferase activity increased to about 2.7% of *Renilla* luciferase activity in R-HRV-F<sub>A30</sub>-transfected cells. As shown in Fig. 4B, the IRES of HRV enhanced the expression of the second cistron by about 40-fold as compared with vector control. However, the 5'-UTR sequence of PDGF-B did not enhance the expression of the second cistron. In fact, it reduced the basal level expression of the second cistron. Thus, there is no detectable IRES activity in the long 5'-UTR sequence of PDGF-B that mediates internal translation initiation.

**The 5'-UTR Promoter Has Two Transcription Start Sites for Production of mRNAs with Shorter 5'-UTRs**—To further determine whether the long 5'-UTR sequence of PDGF-B contains putative promoters, we tested whether there are any transcripts that are derived from the 5'-UTR sequence in cells transfected with the dicistronic constructs. For this purpose, poly(A) RNAs were isolated for Northern blot analysis from cells transfected with pRF, pR-PDGF-F, pR-PDGF-F(-P), and pRF(-R). The pRF(-R) construct, which lacks the *Renilla* lucif-



**FIG. 3. Dicistronic DNA test of the 5'-UTR sequence of human PDGF-B.** *A*, schematic diagram of traditional and promoterless dicistronic constructs. The 5'-UTR sequence of human PDGF-B and the IRES element of HRV are cloned into the intergenic region of the dicistronic vectors. The locations of several relevant restriction enzyme sites are shown by arrows. In the promoterless constructs, the SV40 promoter and the chimeric intron sequence were deleted. *B*, relative luciferase activities generated by the dicistronic constructs in H1299 cells. H1299 cells were transfected with dicistronic constructs together with  $\beta$ -galactosidase plasmid. Twenty-four hours following transfection, cells were harvested and the *Renilla* and firefly luciferase activities were measured and normalized against  $\beta$ -galactosidase activity followed by normalization against the firefly luciferase activity of pRF. The data shown represent one of the four independent experiments.

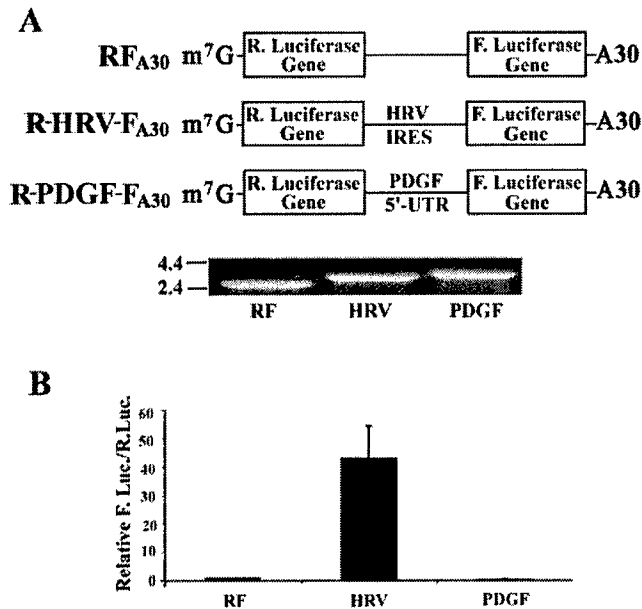
erase gene, was used as a monocistronic control (25). It is expected to use the SV40 promoter to produce a transcript with a 5'-UTR of ~100 bases and the firefly luciferase-encoding sequence. The pRF vector was used as a dicistronic control, which produces only a dicistronic mRNA using the SV40 promoter. As shown in Fig. 5A, a dicistronic transcript from control pRF (lane 2, indicated by an asterisk) and a monocistronic transcript from control pRF(-R) (lane 3, indicated by an arrowhead) were detected as expected. The dicistronic transcript derived from pR-PDGF-F (lane 1, indicated by an asterisk) has a slower mobility than the one from pRF (lane 2), consistent with the presence of the 5'-UTR sequence of PDGF-B in the intergenic region. Two additional transcripts were generated from pR-PDGF-F (lane 1, indicated by A and B), which have similar sizes to the monocistronic transcript derived from pRF(-R) (lane 3), suggesting that they may be monocistronic mRNA transcribed from the 5'-UTR sequence of PDGF-B located in the intergenic region. This conclusion was further confirmed by the production of the same two transcripts from pR-PDGF-F(-P) construct (lane 4), which lacks the vector SV40 promoter and thus the dicistronic product (lane 4). The bigger transcript (labeled as A) may have a longer 5'-UTR than the smaller one (labeled as B). The 5'-UTR sequence of the transcript B may be very short, because it was smaller than the monocistronic mRNA transcript generated from pRF(-R) (compare lanes 4 and 3). Based on the above observations, we conclude that the 5'-UTR sequence of PDGF-B may contain two discrete promoters (designated as P1 and P2) that mediate the production of two types of transcripts with medium and short 5'-UTR sequences, respectively. The transcripts with the short 5'-UTR is likely the one that can be much more efficiently

translated than the one with medium and the original long 5'-UTRs as shown in Fig. 2.

**Delineation of the P1 and P2 Promoter**—To locate the transcription start sites of the P1 and P2 promoter, RT-PCR was performed. Poly(A) RNAs from cells transfected with pR-PDGF-F and pR-PDGF-F(-P) were used as templates for cDNA synthesis using antisense primer complementary to the +60 to +78 sequence of the firefly luciferase gene. The PCR was performed using antisense primer located at the +675 and sense primers located at the +150, +225, +300, and +375 of the 5'-UTR sequence of PDGF-B, respectively. The pR-PDGF-F, which is expected to generate the dicistronic mRNA (see Fig. 3A) that contains the full-length 5'-UTR of PDGF-B, was used to serve as a positive control. As shown in Fig. 5B, although there are similar amounts of PCR products between pR-PDGF-F and pR-PDGF-F(-P) when primers P225, P300, and P375 were used (lanes 4–8), much less PCR product was observed with pR-PDGF(-P) when the primer P150 was used (lane 2), suggesting that the major transcription start site from P1 in pR-PDGF-F(-P) transfection is located downstream of nucleotide +150.

To further delineate the boundaries of the DNA region that are responsible for promoter activities, systematic deletion mutants were created from either 5'- or 3'-end of the 5'-UTR of PDGF-B (Fig. 6A). These 5'-UTRs with deletions were engineered into the promoterless dicistronic vector and used to determine their ability to direct firefly luciferase expression in both H1299 and HeLa cells (Fig. 6B). It should be noted, however, that the results of the luciferase reporter assay may be subject to both transcriptional and translational control when mRNAs with different lengths of 5'-UTRs are produced (27).

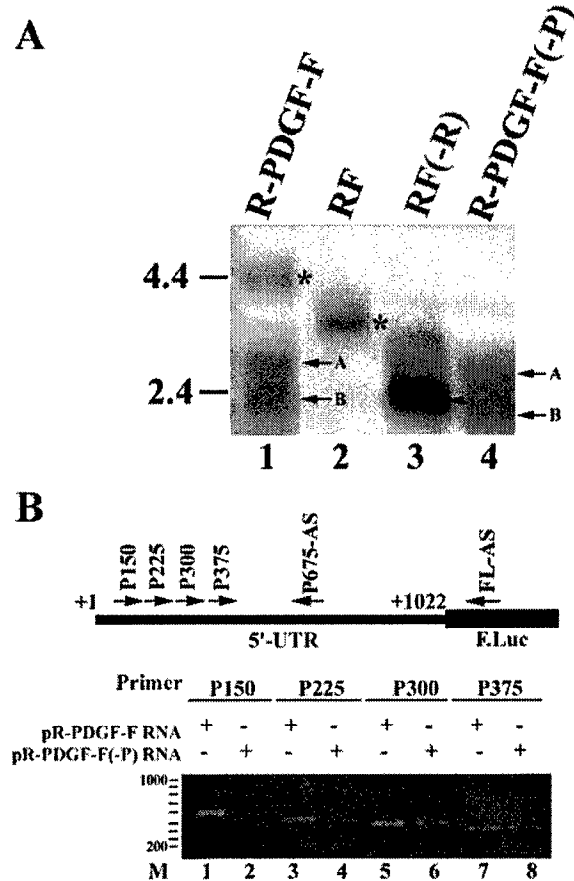




**FIG. 4. Analysis of the IRES activity of the 5'-UTR sequence of PDGF-B in H1299 cells by RNA transfection.** A, schematic diagram and *in vitro* transcripts of the dicistronic mRNA used for translation in H1299 cells. *In vitro* transcripts with 5'-cap ( $m^7G$ ) and 3'-poly(A) tail ( $A_{30}$ ) were synthesized using T7 RNA polymerase from linearized vector ( $RF_{A30}$ ), constructs containing the IRES of HRV ( $R-HRV-F_{A30}$ ), and the PDGF-B 5'-UTR ( $R-PDGF-F_{A30}$ ). The gel shows the profile of the transcripts produced. B, relative luciferase activity from dicistronic mRNAs in H1299 cells. H1299 cells were transfected with the dicistronic mRNAs, and 8 h following transfection, *Renilla* and firefly luciferase activities were measured. The firefly to *Renilla* luciferase ratios were calculated and these values were then divided by that of the vector-transfected cells ( $RF_{A30}$ ) to obtain the relative firefly to *Renilla* luciferase ratios.

Thus, the data in Fig. 6 should be analyzed together with the Northern blot and RT-PCR result shown in Fig. 5B. The data in Fig. 6 clearly demonstrated that two discrete regions, i.e. the +1 to +395 (P1) and the +769 to +1022 (P2) sequence, contain promoter activity. This result confirmed the data in Fig. 5 where we detected two discrete transcripts (Fig. 5A, labeled as A and B in lane 4). The P1 promoter activity is about 3- to 4-fold of that of the P2 promoter. Because transcripts from the P1 promoter have a relatively long 5'-UTR, they can not be efficiently translated (see Fig. 2). The +1 to +769 construct does not contain the P2 sequence. However, because it contains the full P1 promoter, the +1 to +769 construct can produce a transcript with a shorter 5'-UTR that lacks the downstream major GC-rich region. As shown in Fig. 2A, the computed free energy for the +769 to +1022 is about  $-119$  kcal/mol. Removal of this region is thus expected to result in an enhanced translation efficiency and, consequently, an increase in luciferase activity (Fig. 6, B and C). Further deletion from the 3'-end up to 675 bases resulted in further increase in luciferase activity (+675 to +1022). The luciferase activity started to drop in the +1 to +303 construct in both cell lines, indicating that the +303 to +395 region may contain enhancing elements for P1 promoter.

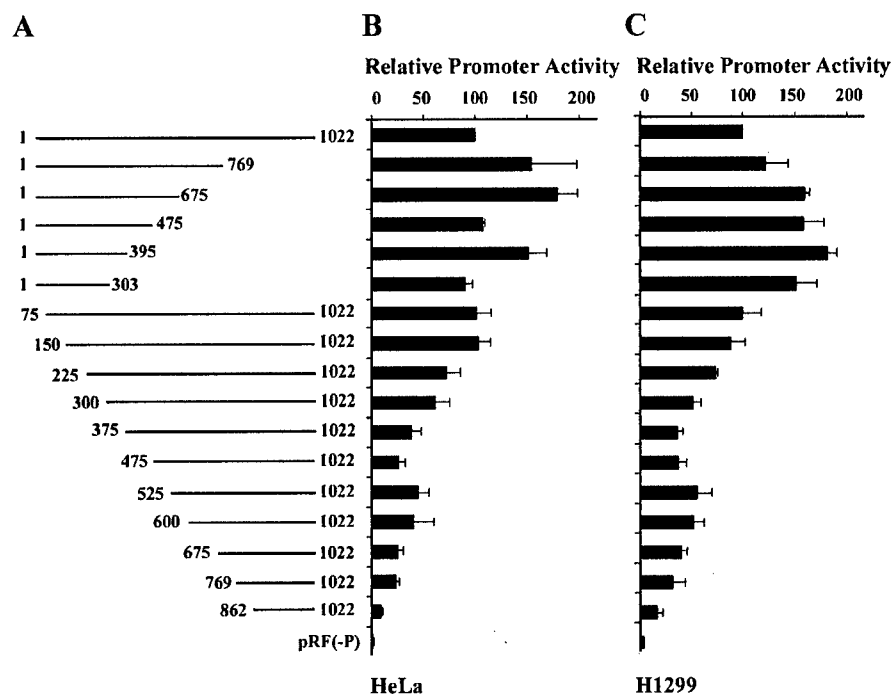
Deletion of 150 bases from the 5'-end did not affect luciferase activity (+150 to +1022) as compared with the +1 to +1022 construct. However, deletion of 225, 300, and 375 bases from the 5'-end significantly decreased the luciferase activity. Combined with the 3'-end deletion construct data, these results indicate that the P1 promoter sequence is located roughly in the +150 to +395 region. The P1 transcript is thus expected to be poorly translated, because it contains a significant secondary structure with an approximate free energy of  $-280$  kcal/



**FIG. 5. Northern blot and RT-PCR analysis of RNA products generated by PDGF-B 5'-UTR promoter.** A, poly(A) RNAs were isolated from H1299 cells following transfection with pR-PDGF-F (lane 1), pRF (lane 2), pRF(-R) (lane 3), pR-PDGF-F(-P) (lane 4), and used for Northern blot analysis as described under "Experimental Procedures." The asterisks indicate the dicistronic mRNA containing both *Renilla* and firefly luciferase. The arrowhead shows the major monocistronic mRNA transcribed from pRF(-R). The arrows with letters A and B indicate two distinct monocistronic transcripts derived from promoters in the 5'-UTR sequence of PDGF-B. B, analysis of the approximate transcription start site of the longer transcript from the 5'-UTR promoter by RT-PCR. The top panel shows the schematic diagram of relative locations of primers used in the RT-PCR. The FL-AS primer is located 78 bases downstream of ATG. PCR amplification was performed using the antisense primer (P675-AS) located at +675 and four sense primers located at +150 (P150) (lanes 1 and 2), +225 (P225) (lanes 3 and 4), +300 (P300) (lanes 5 and 6), and +375 (P375) (lanes 7 and 8), respectively. The bottom panel shows the product of RT-PCR. Total RNAs isolated from pR-PDGF-F and pR-PDGF-F(-P)-transfected H1299 cells were used for RT-PCR analysis. The final DNA products were separated by agarose gel electrophoresis.

mol (Fig. 2). Nevertheless, the integrity of the P1 promoter is important for the activity of the full-length 5'-UTR promoter, because deletion of the P1 promoter from the 5'-UTR (+375 to +1022) significantly decreased the luciferase activity. It is likely that the P1 promoter enhances transcription from the P2 promoter in the intact 5'-UTR construct. The deletion mutants with the 5'-end from +375 to +769 displayed similar luciferase activities, although a slight variation in luciferase activity was observed, arguing for the P2 promoter being located roughly between +769 and +1022. The transcript from the P2 promoter would contain significantly shorter 5'-UTRs than that from the P1 promoter. This is consistent with our observation of the transcript with a short 5'-UTR in Northern blot analysis (labeled as B in Fig. 5A, lane 4). The P2 promoter is likely responsible for production of the endogenous 2.8-kb mRNA with short 5'-UTRs of 15-27 bases.

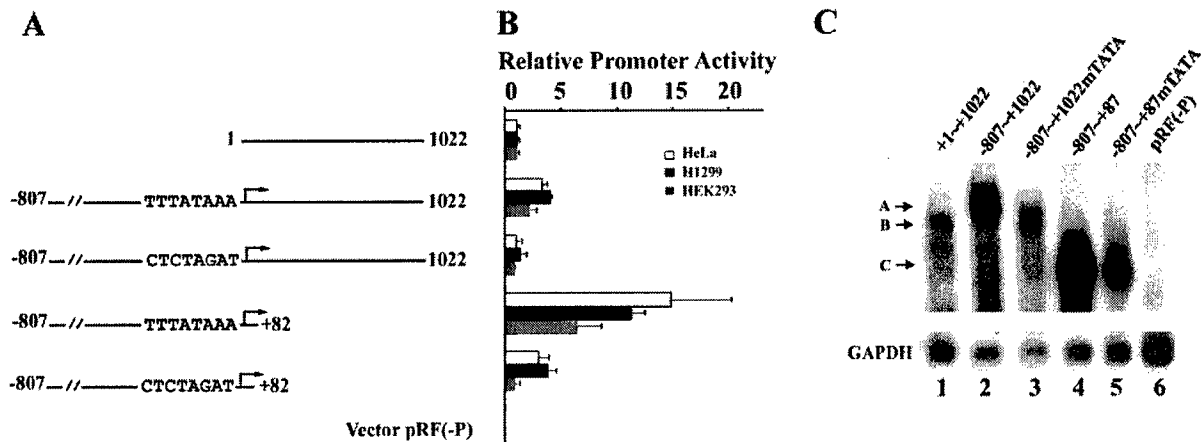
**FIG. 6. Deletion mapping of the 5'-UTR promoter of PDGF-B.** A, schematic diagram of the deletion constructs of PDGF-B 5'-UTR. The positions of the 5'- and 3'-ends of each deletion are indicated on the left and right, respectively. These mutant 5'-UTRs were engineered into the promoterless dicistronic vector pRF(-P) at the intergenic region. B and C, relative luciferase activity from H1229 (B) and HeLa (C) cells transfected with the wild type and mutant PDGF-B 5'-UTR constructs. Twenty-four hours following transfection, cells were harvested for *Renilla* and firefly luciferase activity determination. The ratio of firefly to *Renilla* luciferase was calculated and normalized to the full-length 5'-UTR construct. The data were from three independent experiments.



**Function of the 5'-UTR Promoter in the Presence of the Upstream TATA-containing Promoter (P0)**—The above promoter analysis focused on the 5'-UTR sequences only. However, the promoter activity of the 5'-UTR sequence may be different in the presence of the upstream TATA-containing P0 promoter. To address this issue, we generated a series of constructs as shown in Fig. 7A and analyzed the promoter activity using both luciferase assay (Fig. 7B) and Northern blot (Fig. 7C). The longest construct, from -807 to +1022, contained both the TATA-containing promoter (P0) and the full-length 5'-UTR sequence. It has been known that the TATA-box is critical for constitutive promoter activity for P0 promoter and, thus, a TATA-box mutant construct (from -807 to +1022 mTATA with mutation from TTTATAAAA to CTCTAGAT, which has no homology with any known transcription factor binding sites) was also generated to determine the effect of P0 promoter on expression from the -807 to +1022 sequence. Two other constructs (those from -807 to +87 and -807 to +87 mTATA) were also generated to determine the P0 activity in the absence of the 5'-UTR. These plasmids were transfected into HeLa, H1229, and HEK293 cells for promoter analysis. As shown in Fig. 7B, the -807 to +1022 construct has about a 3-, 4-, and 2-fold increase in luciferase expression in HeLa, H1229, and HEK293 cells, respectively, as compared with the control +1 to +1022 construct (pR-PDGF-F(-P)) that has only the 5'-UTR sequence. Mutation of the TATA-box almost eliminated the enhancement, suggesting that the integrity of the P0 promoter affects transcription from the -807 to +1022 construct. On the other hand, removal of the 5'-UTR sequence (-807 to +87 construct) enhanced the reporter expression about 4- to 5-fold compared with the full-length construct (-807 to +1022), indicating that the 5'-UTR caused an inhibitory effect on luciferase expression in the -807 to +1022 construct. This observation is consistent with the data shown in Fig. 2B. The -807 to +87mTATA construct had about 20-30% of luciferase activity compared with the -807 to +87 construct, which again demonstrate that the TATA-box is critical for the activity of the P0 promoter.

The constructs in Fig. 7A are expected to produce mRNAs with different lengths of 5'-UTRs and, thus, likely affect the

results of the luciferase reporter assay (Fig. 7B) due to the possible effect of 5'-UTR sequence on translation of the reporter. To address this issue, we performed Northern blot analysis. As shown in Fig. 7C, cells transfected with the +1 to +1022 construct but not the vector pRF(-P) control produced two transcripts labeled B and C (compare lanes 1 and 6), derived from P1 and P2 promoters, respectively. The -807 to +1022 construct produced both a major product labeled A and the other minor diffused band labeled C (lane 2). Transcript A is the largest transcript derived from P0 promoter and has a 5'-UTR longer than that of the transcript C, which is derived from promoter P2. Interestingly, the -807 to +1022 mTATA-transfected cells generated three distinct bands labeled as A, B, and C, derived from P0, P1, and P2, respectively (lane 3). It appears that the TATA-box mutation dramatically reduced the transcription from P0 promoter and, thus, reduced production of the transcript A from the P0 promoter. Such an effect was also observed with the TATA-box mutation in the -807 to +87 construct (compare lane 4 and lane 5). However, the transcript B derived from the P1 promoter was not produced from the -807 to +1022 construct (lane 2). It is possible that the P0 and P1 promoter are integrated into one promoter due to their nearness to each other. Because the P0 promoter is about 10-fold more active than the P1 promoter, it is possible that the stronger promoter (P0) may dominate the weaker one (P1). This is consistent with the observation that the reduction of the P0 promoter activity by TATA-box mutation is accompanied by the appearance of transcript B from P1 promoter at a similar level to that produced from the +1 to +1022 construct (compare lanes 1 and 3). Although the level of transcript A produced from the -807 to +1022 construct is much higher, the luciferase activity produced is only about 2- to 4-fold over that of the +1 to +1022 construct (Fig. 7B). Furthermore, despite the production of the additional transcript A in cells transfected with the -807 to +1022 mTATA construct, the luciferase activity in these cells is almost equal to that of cells transfected with the +1 to +1022 construct (see Fig. 7B). Thus, we conclude that the transcript A produced from the -807 to +1022 and the -807 to +1022 mTATA constructs contributes little if any to the luciferase activity. The fact that the transcription



**FIG. 7. Function of the 5'-UTR promoters of PDGF-B in the presence or absence of the major TATA-containing promoter.** **A**, schematic diagram of the constructs with or without the TATA-containing promoter. The positions of the 5'- and 3'-ends of each deletion are indicated on the left and right, respectively. Also shown are the constructs with mutated TATA box. All constructs were engineered with the promoterless dicistronic vector pRF(-P). **B**, relative luciferase activity from H1299, HeLa, and HEK293 cells transfected by different constructs. Twenty-four hours following transfection, cells were harvested for *Renilla* and firefly luciferase activity determination. The ratio of firefly to *Renilla* luciferase was calculated and normalized to the full-length 5'-UTR construct. The data were from three independent experiments. **C**, Northern blot analysis. Poly(A) RNAs were isolated from H1299 cells transfected with the constructs shown in **A** and used for Northern blot analysis as described under "Experimental Procedures." The arrows labeled with A-C indicate the RNAs derived from P0, P1, and P2 promoters, respectively. Glyceraldehyde-3-phosphate dehydrogenase was used as a control.

from the P1 promoter occurs only upon significant reduction of the P0 promoter activity suggests that a competition of transcription between the P0 and P1 promoter may exist. Thus, it is possible that the P1 promoter does not function *in vivo* in the presence of the P0 promoter. This possibility is consistent with our observation that only the 3.8- and 2.8-kb endogenous PDGF-B mRNAs were detected.

**Induction of the P0, P1, and P2 Promoter during Megakaryocytic Differentiation.**—We next examined the induced activity of the 5'-UTR promoter and the major P0 promoter in K562 cells upon TPA stimulation. For this purpose, K562 cells were transiently transfected with various constructs (Fig. 8A) by electroporation and were then treated without (Fig. 8B) or with (Fig. 8C) TPA for 2 days followed by measuring luciferase activities in cell lysates. The vector control was used to normalize the luciferase activities. As shown in Fig. 8B, the 5'-UTR sequence (+1 to +1022) alone stimulated about 10-fold expression compared with the vector control in K562 cells. In the presence of both P0 promoter and the 5'-UTR sequence (-807 to +1022), the activity increased another 2.6-fold. However, mutation of the P0 promoter eliminated this increase (-807 to +1022 mTATA). Deletion of the 5'-UTR sequence dramatically increased the expression of luciferase in K562 cells (~11-fold increase) (compare the -807 to +82 range with that from -807 to +1022), suggesting that the long 5'-UTR sequence inhibits translation. These observations are consistent with the results derived from 293, HeLa, and H1299 cells shown in Fig. 7.

Upon TPA-induced megakaryocytic differentiation, the 5'-UTR promoter (+1 to +1022) activity increased ~2.7-fold (Fig. 8C). Interestingly, this result is in agreement with a 2–3-fold increase of "IRES" activity by induction observed previously (29), suggesting that the promoter activity of the 5'-UTR sequence might be misinterpreted as IRES activity. In the presence of the P0 promoter, however, the production of the luciferase reporter increased ~10-fold by TPA stimulation (compare -807 to +1022 with +1 to +1022). The induction of luciferase expression of P0 promoter without the 5'-UTR sequence was ~20-fold (-807 to +87), demonstrating that the P0 promoter is the major promoter that responds to TPA stimulation as compared with the 5'-UTR promoter.

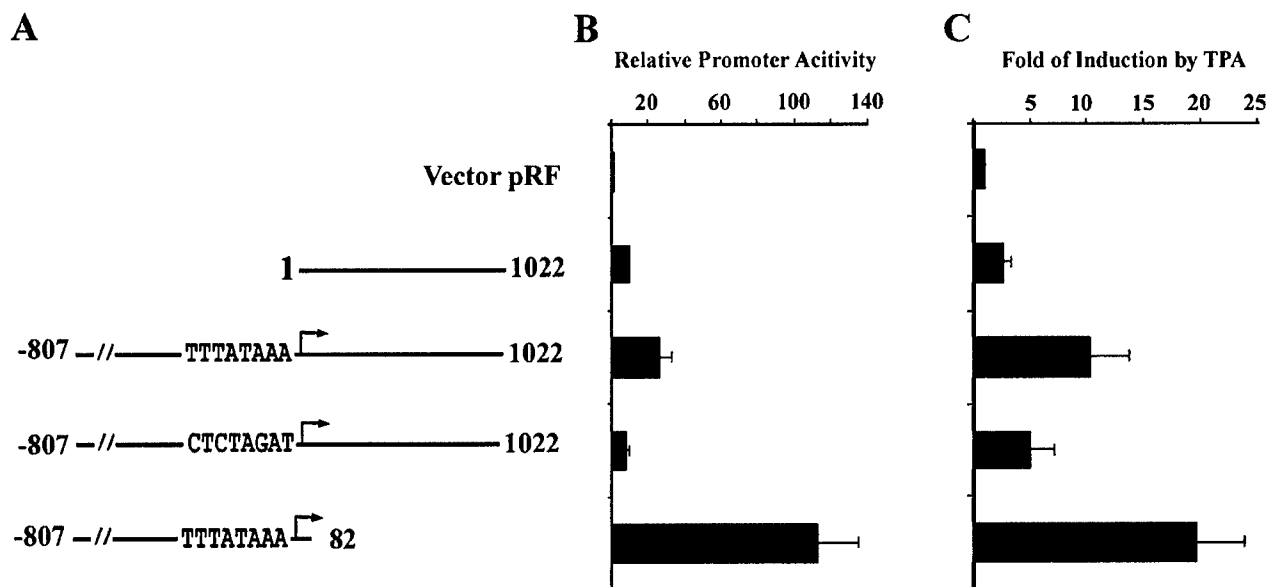
The luciferase activity from the construct containing P0 and 5'-UTR (the P1 and P2 promoter) increased about 10-fold,

which is higher than the increase observed with the 5'-UTR promoter (P1 and P2) alone (2.7-fold) but lower than that by the P0 promoter alone (20-fold). Disruption of the integrity of the upstream promoter by mutation of the TATA box significantly decreased the production of luciferase protein from 20- to 5-fold (Fig. 8C). However, the -fold increase in luciferase activity with the construct containing mutated TATA box was still higher than that with the 5'-UTR sequence alone. This is consistent with results previously reported by others (6, 16) that cis-elements other than TATA signals in the P0 promoter also contribute to TPA induction. Because the transcript derived from the P0 promoter contains the full-length 5'-UTR that inhibits translation and only the transcripts derived from P2 can be efficiently translated, we propose that the transcription from the 5'-UTR promoter is greatly enhanced in the presence of the upstream promoter by TPA stimulation.

#### DISCUSSION

In this study, using firefly luciferase as a heterologous reporter gene, we found that the full-length 5'-UTR of PDGF-B inhibits translation of the luciferase gene in a rabbit reticulocyte lysate cell-free system (Fig. 2). We also found that the mRNA lacking 5'-UTRs or containing shorter 5'-UTRs are 30- to 70-fold more efficiently translated. These observations are consistent with previous studies, which showed that the long 5'-UTR inhibits the translation of PDGF-B mRNA (4, 28). However, we found no evidence that the long 5'-UTR of PDGF-B contains any IRES activity that has been suggested to bypass the observed translational inhibition of PDGF-B mRNA (17, 18). Instead, we found that the long 5'-UTR sequence of PDGF-B contains two promoters (P1 and P2) that can produce mRNAs with shorter 5'-UTRs. These two promoters were demonstrated to work coordinately with the upstream TATA-box containing promoter.

By using RNA transfection, we demonstrated that dicistronic mRNAs bearing a full-length 5'-UTR of PDGF-B in the intergenic region does not enhance the translation of the downstream cistron (Fig. 4). Although in a conventional dicistronic DNA transfection assays, cells transfected with a dicistronic construct, which contains the full-length 5'-UTR of PDGF-B in the intergenic region, showed an enhanced expression of the second cistron, no significant decrease in the expression of the



**FIG. 8. Promoter analysis in K562 cells upon differentiation.** A, schematic diagrams of the constructs with or without the TATA-containing promoter the same as shown in Fig. 7A. B, relative promoter activity from K562 cells transfected with different constructs. Cells were harvested 48 h following transfection by electroporation and the firefly luciferase activity was measured and was normalized against that of the pRF(-P) vector control. C, -fold induction of promoter activity by TPA in K562 cells. Twenty-four hours following transfection by electroporation, cells were treated with or without 2 ng/ml TPA for 48 h and collected for determination of luciferase activity. The -fold induction was calculated by dividing the luciferase activity of cells treated with TPA by that of cells not treated with TPA. The data are derived from four independent experiments.

second cistron was observed after the SV40 promoter of the vector was removed (Fig. 3). These observations suggest that the 5'-UTR sequence of PDGF-B contains promoters that underlie the enhanced expression of the second cistron observed in the traditional dicistronic DNA assay. We also found that the 5'-UTR promoter activity increased by about 2- to 3-fold in K562 cells undergoing megakaryocytic differentiation following TPA treatment (Fig. 8). This increase is surprisingly similar to the reported differentiation-induced IRES activity (18, 29). Thus, it is likely that the previously claimed IRES activity of PDGF-B was misinterpreted from the promoter activity present in the 5'-UTR sequence of PDGF-B. In light of these findings, we propose that the effective production of PDGF-B protein mainly relies on the production of the transcripts with shorter 5'-UTRs generated from transcription using the 5'-UTR promoter.

Although the existence of promoter activities in the 5'-UTR sequence of PDGF-B has been proposed previously (10, 12, 13), we, for the first time, characterized the 5'-UTR promoter and demonstrated that they are inducible and may be responsible for the previously thought IRES activity. The 5'-UTR promoter activity showed about 5-10% of the activity of the major TATA-box containing P0 promoter in various cell lines tested when analyzed by the luciferase reporter promoter assay (Fig. 7B). As shown by Northern blot analysis, the 5'-UTR promoter of PDGF-B can by itself initiate transcription from discrete locations, with one located around 800 bases upstream (Fig. 5B), and the other is very close to the translation start codon (Figs. 5 and 7). Detailed promoter analysis revealed two distinctive promoter regions, located approximately in position +150 to +303 (P1) and +769 to +1022 (P2) of the 5'-UTR sequence, respectively (Fig. 6). The P1 promoter in the 5'-UTR is about 3- to 4-fold more active than P2 in both HeLa and H1299 cells (Fig. 6).

We further investigated the function of the 5'-UTR promoter in the presence of the major TATA-box containing promoter (P0) (Fig. 7). TATA-box has previously been shown to be critical for proper function of the P0 promoter activity (5). This is confirmed in our promoter assay using both reporter luciferase

activity assay and Northern blot analysis (compare the result of the -807 to +87 construct with that of the -807 to +87 mTATA construct). Although the major promoter P0 alone is about 7- to 20-fold more active than the 5'-UTR promoter (compare the result of the -807 to +87 construct with that of the +1 to +1022 construct), the luciferase activity increased only 2- to 4-fold when they coexist. Mutation of the TATA-box in the -807 to +1022 construct caused the luciferase activity to drop to the similar level to that of the +1 to +1022 construct. Although the P2 promoter functions in all of the constructs, the P1 promoter functions only in the absence of the P0 promoter or in the presence of the P0 promoter with a mutated TATA box (see Fig. 7C). Furthermore, the transcript generated from the P0 promoter with the -807 to +1022mTATA construct did not cause any increase in the luciferase activity produced. Thus, it is likely that the P0 and P1 promoter are integrated into one promoter and drives transcription of the 3.8-kb full-length transcript that can be poorly translated *in vivo*.

We also analyzed the induction of the promoter activities in K562 cells upon TPA-induced megakaryocytic differentiation (Fig. 8) and demonstrated that the P0 promoter is a major player to respond to TPA stimulation, whereas the P1 and P2 promoters in the absence of P0 are induced only in a minor fashion. However, when P0 and the 5'-UTR promoter coexist, the overall protein production increased about 10-fold upon induction in transient transfection assays. Stable clones transfected with the -807 to +1022 construct with all three promoters displayed a similar induction profile.<sup>2</sup> Based on these observations, we propose that transcriptions from the 5'-UTR promoters are greatly induced by TPA in the presence of the P0 promoter, and P0 and the 5'-UTR promoters work in a highly coordinated manner to tightly control the production of transcripts both with the long and short 5'-UTRs and, therefore, the constitutive and induced production of PDGF-B protein. Although cis-elements and transcription factors has been well characterized for the P0 promoter (5, 7, 15, 16, 30, 31), further

<sup>2</sup> B. Han, Z. Dong, and J.-T. Zhang, unpublished observation.

studies is clearly needed to characterize the 5'-UTR promoter. Nevertheless, our experiments using *Drosophila* SL2 cells transfected with 5'-UTR promoter constructs demonstrated that Sp1 might be one of the transcription factors that can trans-activate the 5'-UTR promoter.<sup>2</sup>

Our promoter analysis strongly supports the argument that the 2.8-kb PDGF-B mRNA with short 5'-UTR is likely derived from transcription from the P2 promoter of the 5'-UTR sequence as previously suggested (10, 12, 13), although there is no evidence available for the function of the natural P2 promoter of PDGF-B on human chromosome 22. The integrated P0/P1 promoter likely drives the production of the 3.8-kb full-length PDGF-B mRNAs. The finding that the dramatic increase of 2.8-kb mRNA species in K562 cells following TPA treatment (Fig. 1A) contradicts with only a 2-fold induction of the 5'-UTR promoter activity in the absence of P0 (Fig. 8C) and argues for a coordination between the P0/P1 promoter and the P2 promoter.

In the analysis of PDGF-B mRNA in K562 cells upon TPA-induced differentiation, we clearly detected a 2.8-kb mRNA that has a 5'-UTR of about 15–27 bases. The 2.8-kb mRNA was not reported for K562 cells in previous studies, possibly because it represents only a minor species and is not evident in Northern blot with short exposures. We, however, noticed that a Northern blot performed by Colamonici *et al.* (14) clearly showed a 2.8-kb PDGF-B mRNA in K562 cells that were treated for 3 or 4 days (see Fig. 2 in Ref. 14), although it is not discussed in this report. Similarly, Fen and Daniel (13) reported the detection of a 2.8-kb mRNA species in human renal microvascular endothelial cells (HRMECs) upon treatment with either TGF- $\beta$  or TPA. The 5'-truncated mRNAs were also detected in a few tumor cells and in rat brain tissue at a certain stage of development, and its level correlates the level of PDGF-B protein (12). Based on these observations, we propose that the production of the 2.8-kb mRNA may be tightly regulated and widely used for effective protein production of PDGF-B, both constitutively and upon induction by biological stimuli. However, it remains unknown what is the role of the 3.8-kb mRNA of PDGF-B if it cannot be used as template for efficient protein synthesis.

Consistent with Fen and Daniel (13), the 2.8-kb mRNA in TPA-induced K562 cells is selectively enriched by cycloheximide treatment. Although cycloheximide likely does not affect the transcription of PDGF-B (see Fig. 1, lanes 1 and 2), we cannot rule out the possibility that cycloheximide treatment may inhibit the synthesis of a suppressor for transcription and thus, results in the increase in the 2.8-kb mRNA. However, the analysis of PDGF-B mRNA half-life by Fen and Daniel demonstrated that the 2.8-kb mRNA undergoes rapid decay without cycloheximide treatment, arguing for that cycloheximide treatment is associated with decay of the 2.8-kb mRNA. On the other hand, the 2.8-kb mRNA is more abundant in some tumor cells and in developing rat brain tissues at a certain stage (10, 12). Such abundance may likely be a result of an increase in either transcription from the P2 promoter or alternatively an increase in mRNA stability. The regulation of mRNA stability has been demonstrated to play a major role in regulating the expression level of oncogene and cytokine mRNAs during cell growth, differentiation, and neoplastic transformation (32). The mRNA decay rates are regulated by cis-acting sequence determinants, mRNA-binding proteins, endo- and exo-ribonucleases, and translation (32). For example, *c-myc* encoding

region contains a coding region instability determinant (CRD), located in the last 249 nucleotides of the coding region, which mediates rapid turnover of *c-myc* mRNA (33). The CRD-mediated turnover is translation-dependent and blocked by translation inhibitor cycloheximide. The CRD-BP protein specifically binds to the CRD region and prevented *c-myc* mRNA from degradation (34). CRD-BP expression parallels *c-myc* expression during liver development. Interestingly, the abundance of the 2.8-kb mRNA of PDGF-B is also developmentally regulated in rat brain (12). Because the level of the 2.8-kb mRNA species is a major determinant of PDGF-B protein level (10, 12), alteration of the stability may be another level of control in PDGF-B expression regulation. How the stability of the 2.8-kb mRNA is regulated remains to be an intriguing question for further studies.

**Acknowledgments**—We are indebted to Dr. A. E. Willis for pRF and pGL3RHRV plasmids and Dr. William E. Fahl for the plasmid pBLH-HCAT. We also thank Aaron Smith for his effort in proof reading of the manuscript.

#### REFERENCES

- Betsholtz, C., Karlsson, L., and Lindahl, P. (2001) *Bioessays* **23**, 494–507
- Heldin, C. H., and Westermark, B. (1999) *Physiol. Rev.* **79**, 1283–1316
- Mariani, S., Basciani, S., Arizzi, M., Spera, C., and Gnassi, L. (2002) *Trends Endocrinol. Metab.* **13**, 11–17
- Ratner, L., Thielan, B., and Collins, T. (1987) *Nucleic Acids Res.* **15**, 6017–6036
- Jin, H. M., Robinson, D. F., Liang, Y., and Fahl, W. E. (1994) *J. Biol. Chem.* **269**, 28648–28654
- Pech, M., Rao, C. D., Robbins, K. C., and Aaronson, S. A. (1989) *Mol. Cell. Biol.* **9**, 396–405
- Khachigian, L. M., Fries, J. W., Benz, M. W., Bonthron, D. T., and Collins, T. (1994) *J. Biol. Chem.* **269**, 22647–22656
- Betsholtz, C., Johnsson, A., Heldin, C. H., Westermark, B., Lind, P., Urdea, M. S., Eddy, R., Shows, T. B., Philpott, K., and Mellor, A. L., Knott, T. J., and Scott, J. (1986) *Nature* **320**, 695–699
- Igarashi, H., Rao, C. D., Siroff, M., Leal, F., Robbins, K. C., and Aaronson, S. A. (1987) *Oncogene* **1**, 79–85
- Heller, S., Scheibenpflug, L., Westermark, B., and Nister, M. (2000) *Int. J. Cancer* **85**, 211–222
- Wang, J., Coltrera, M. D., and Gown, A. M. (1994) *Cancer Res.* **54**, 560–564
- Sasahara, M., Amano, S., Sato, H., Yang, J. G., Hayase, Y., Kaneko, M., Sato, I., Suzuki, M., and Hazama, F. (1998) *Oncogene* **16**, 1571–1578
- Fen, Z., and Daniel, T. O. (1991) *Oncogene* **6**, 953–959
- Colamonici, O. R., Trepel, J. B., Vidal, C. A., and Neckers, L. M. (1986) *Mol. Cell. Biol.* **6**, 1847–1850
- Jin, H. M., Brady, M. L., and Fahl, W. E. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7563–7567
- Kujoth, G. C., Robinson, D. F., and Fahl, W. E. (1998) *Cell Growth & Differ.* **9**, 523–534
- Sella, O., Gerlitz, G., Le, S. Y., and Elroy-Stein, O. (1999) *Mol. Cell. Biol.* **19**, 5429–5440
- Bernstein, J., Sella, O., Le, S. Y., and Elroy-Stein, O. (1997) *J. Biol. Chem.* **272**, 9356–9362
- Hellen, C. U., and Sarnow, P. (2001) *Genes Dev.* **15**, 1593–1612
- Sachs, A. B. (2000) *Cell* **101**, 243–245
- Vagner, S., Galy, B., and Pyronnet, S. (2001) *EMBO Rep.* **2**, 893–898
- Kozak, M. (2001) *Mol. Cell. Biol.* **21**, 1899–1907
- Schneider, R., and Kozak, M. (2001) *Mol. Cell. Biol.* **21**, 8238–8246
- Clarke, M. F., Westin, E., Schmidt, D., Josephs, S. F., Ratner, L., Wong-Staal, F., Gallo, R. C., and Reitz, M. S., Jr. (1984) *Nature* **308**, 464–467
- Han, B., and Zhang, J. T. (2002) *Mol. Cell. Biol.* **22**, 7372–7384
- Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol.* **288**, 911–940
- Han, B., Dong, Z., Liu, Y., Chen, Q., Hashimoto, K., and Zhang, J. T. (2003) *Oncogene* **22**, 5325–5337
- Rao, C. D., Pech, M., Robbins, K. C., and Aaronson, S. A. (1988) *Mol. Cell. Biol.* **8**, 284–292
- Gerlitz, G., Jagus, R., and Elroy-Stein, O. (2002) *Eur. J. Biochem.* **269**, 2810–2819
- Taylor, L. M., and Khachigian, L. M. (2000) *J. Biol. Chem.* **275**, 16709–16716
- Khachigian, L. M., Lindner, V., Williams, A. J., and Collins, T. (1996) *Science* **271**, 1427–1431
- Ross, J. (1995) *Microbiol. Rev.* **59**, 423–450
- Wisdom, R., and Lee, W. (1991) *Genes Dev.* **5**, 232–243
- Leeds, P., Kren, B. T., Boylan, J. M., Betz, N. A., Steer, C. J., Gruppiso, P. A., and Ross, J. (1997) *Oncogene* **14**, 1279–1286
- Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) *Nucleic Acids Res.* **23**, 4878–4884

**Regulation of expression by promoters vs internal ribosome entry site in the 5'-untranslated sequence of the human cyclin-dependent kinase inhibitor p27<sup>kip1</sup>**

Zhaoqian Liu<sup>\*§</sup>, Zizheng Dong<sup>\*</sup>, Baoguang Han, Youyun Yang, Yang Liu, and Jian-Ting Zhang<sup>#</sup>

Department of Pharmacology and Toxicology, Indiana University Cancer Center, Walther Oncology Center/Walther Cancer Institute, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA

Running title: Regulation of p27 expression

Keywords: P27<sup>kip1</sup>; promoter; IRES; 5'-UTR; dicistronic

\* These authors contributed equally to this work

§ Current address: Institute of Clinical Pharmacology, Central South University Xiangya School of Medicine, Changsha, P. R. China 410078

# To whom correspondence should be addressed: Department of Pharmacology and Toxicology, Indiana University School of Medicine, 1044 W. Walnut Street, Indianapolis, IN 46202 USA  
Phone: (317)-278-4503, Fax: (317)-274-8046, Email: jianzhan@iupui.edu

This work was supported in part by the National Institutes of Health grant CA94961 and by the Department of Defense grant DAMD17-02-1-0073.

## SUMMARY

P27<sup>kip1</sup> regulates cell proliferation by binding to and inhibiting the activity of cyclin-dependent kinases and its expression oscillates with cell cycle. Recently, it has been suggested from studies using the traditional dicistronic DNA assay that the expression of p27<sup>kip1</sup> is regulated by IRES (internal ribosome entry site)-mediated translation initiation. Considering the inevitable drawbacks of the dicistronic DNA assay which could mislead a promoter activity or alternative splicing to IRES as previously demonstrated, we decided to reanalyze the 5'-UTR sequence of p27<sup>kip1</sup> and tested whether it contains an IRES element or a promoter using more stringent methods such as dicistronic RNA and promoterless dicistronic and monocistronic DNA assays. We found that the 5'-UTR sequence of human p27<sup>kip1</sup> does not have any significant IRES activity. The previously observed IRES activities are likely generated from the promoter activities present in the 5'-UTR sequences of p27<sup>kip1</sup>. The findings in this study indicate that transcriptional regulation likely plays an important role in p27<sup>kip1</sup> expression and confirms that more stringent studies such as promoterless dicistronic and monocistronic DNA and dicistronic RNA tests are required to safeguard any claims of cellular IRES.

## INTRODUCTION

P27<sup>kip1</sup> (referred as p27 in the remaining text) is an inhibitory protein of cyclin-dependent kinases (Cdk) and it plays a key role in regulating Cdk activity during cell cycle progression and in growth arrest (1,2). The expression level of p27 is a critical determinant for the decision of cells in G1 to either withdraw from or commit to the cell cycle and enter S phase. The expression of p27 oscillates with cell cycle and is higher in G1 and lower in other phases (3).

The regulation of p27 expression is precise and has been suggested to occur at many levels including translation initiation (4-6). In 1996, Agrawal et al. (4) and Hengst and Reed (5) found that the expression of p27 may be regulated at the translational level. Millard et al. (7) found that the translation of p27 mRNA in both basal (proliferation) and induced (nonproliferation) states requires a U-rich sequence in the 5'-untranslated region (5'-UTR). Using a dicistronic DNA vector, Miskimins et al. (8) found that an IRES (internal ribosome entry site) element present in the 217 bases of the 5'-UTR of mRNAs may be involved in the translational regulation of murine p27. Using similar strategies, the IRES in the 5'-UTR of human p27, however, was thought to be longer and it may consist of 356 bases (9). Recently, it was shown that an upstream open reading frame in the 5'-UTR of human p27 is important for the IRES activity (10).

Initiation of translation of most eukaryotic mRNAs normally depends on the 5' m<sup>7</sup>GpppN cap structure of mRNAs, which recruits 43S ribosome preinitiation complex via interaction with the cap binding protein eIF4E (11). The translation machinery then migrates downstream until it meets the first AUG codon in the optimal context for initiation of translation (12,13). This scanning model implies that any mRNAs with long 5'-UTRs and complex secondary structures may not be translated efficiently. An alternative mechanism, IRES-mediated translation initiation



has thus been proposed. With this mechanism, ribosomes can be directly recruited to an internal site in a 5'-UTR sequence proximal to the AUG start codon for initiation without the use of the 5'-cap structure. Up to date, many eukaryotic cellular mRNAs have been suggested to contain such IRES activities (see <http://www.rangueil.inserm.fr/IRESdatabase>). While it is possible that some cellular mRNAs may use the IRES mechanism to initiate translation (14), their existence has been challenged due to the inevitable drawbacks of the conventional dicistronic DNA assay used in the majority of these previous studies (15-17). Despite such concerns, many new studies claiming cellular IRES are being published without any rigorous analysis to rule out the potential interference of promoters and alternative splicing events.

In the present study, we re-examined the 5'-UTR sequence of human p27 mRNA and further tested the putative IRES activities using promoterless dicistronic and monocistronic vectors and direct assays of dicistronic mRNAs. Using these stringent assays, we found that the 5'-UTR sequence of human p27 does not have IRES activities but contains promoters which may be involved in regulating p27 expression. These promoter activities are likely responsible for the previously reported putative IRES activities of p27.

## **MATERIALS and METHODS**

**Materials.** Restriction endonucleases and Pfu polymerase were purchased from New England Biolabs and Stratagene, respectively. Sp6 and T7 RNA polymerases, RNasin, RNase-free DNase, Luciferase Reporter Assay Systems, pGEM-T easy and pSP64 polyA vectors were from Promega. RNeasy kit and Oligotex mRNA Mini kit were from Qiagen. MAXIscrip*t in vitro* transcription kit and RPA III Ribonuclease Protection Assay Kit were from Ambion. m<sup>7</sup>GpppG cap analogue and  $\alpha$ -[<sup>32</sup>P]dCTP were from Amersham/Pharmacia. The Sephadex G-25 Quick Spin Column (TE) for radio-labeled DNA and RNA purification was from Roche Diagnostics.

PCR Cloning kit, cell culture medium and serum, Lipofectamine Plus and Lipofectin transfection reagents were all from Invitrogen. All other reagents were of molecular biology grade from Fisher or Sigma.

**Construct engineering.** The full length sequence of 5'-UTR of human p27 is 575 bases (accession numbers: U77915 and U10440). The full-length and the two deletion mutants of 5'-UTR sequences of human p27 were amplified from human genomic DNA using a common reverse primer 5'-GGCCATGGCTTTCTCCCGGGTCTGCACGA-3' and following individual forward primers: 5'-GGACTAGTCCACCTTAAGGCCGCGCTCGC-3' for full length (-575 to -1) 5'-UTR; 5'-GGACTAGTAGCCTCCCTTCCACCGCCATA-3' for deletion construct (-461 to -1); and 5'-GGACTAGTGGCCGTGGCTCGTCGGGGTCT-3' for deletion construct (-150 to -1). These PCR products were then cloned into pGEM-T Easy vector.

To engineer dicistronic constructs containing different 5'-UTRs of human p27, the dicistronic vector pRF that contains *Renilla* and firefly luciferase genes (18) was used. The full length and the two deletion constructs of human p27 5'-UTR in pGEM-T Easy were relieved by digestion with *Spe* I and *Nco* I and subsequently cloned into pRF vector to obtain pR-p27-F(-575), pR-p27-F(-461), and pR-p27-F(-150). The promoterless dicistronic constructs containing these 5'-UTRs were engineered by removing the vector SV40 promoter together with the chimeric intron as previously described (19). The hairpin structure in the dicistronic constructs were engineered by cloning the 5'-UTR sequences into the dicistronic constructs containing a hairpin prepared in a previous study (20). To generate constructs that can be used for *in vitro* production of transcripts with polyA tail, the 5'-UTR sequences in the dicistronic constructs were released by digestion with *Spe* I and *Nco* I and subsequently cloned into the pSP-R-HRV-F<sub>A30</sub> plasmid (19) by replacing the HRV IRES sequence digested with *Spe* I and *Nco* I.

***In vitro* transcription.** *In vitro* transcription was performed as previously described (21). Briefly, pRFA<sub>30</sub>-based plasmids were linearized by *EcoR* I and the capped transcripts were synthesized *in vitro* in the presence of 1 mM m<sup>7</sup>GpppG using Sp6 RNA polymerase. The DNA template was then removed by digestion using RNase-free DNase I and the RNA transcripts were purified using a Qiagen RNeasy Mini kit.

**Cell culture and transfections.** Human large cell lung cancer cell line H1299 was maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum in humidified incubators at 37°C with 5% CO<sub>2</sub>.

Transient transfection of H1299 cells with DNA constructs were performed with Lipofectamine plus reagents as previously described (19,20,22). Briefly, about 6×10<sup>4</sup> cells/well in a 24-well plate were co-transfected with 0.4 µg of test DNA and 50 ng β-galactosidase control DNA. Cells were then harvested 24 hrs post-transfection followed by analysis of luciferase and galactosidase activities.

Transient transfection with RNA transcripts were performed using cationic liposomes also as previously described (19,20,22). Briefly, approximately 2×10<sup>5</sup> cells/well were seeded onto 6-well plates on the day before transfection. Cells were washed once with Opti-MEM I reduced serum medium (Invitrogen) and left in the incubator with some medium during preparation of the liposome-polynucleotide complexes. One ml of Opti-MEM I medium in a 12 × 75 mm polystyrene snap-cap tube was mixed with 12.5 µg of Lipofectin reagent and 5 µg RNA transcripts. The liposome-RNA-medium mixture was immediately added to cells. Eight hours following transfection, cells were harvested and processed for luciferase assay.

**Luciferase reporter assay.** H1299 cells co-transfected with dicistronic or monocistronic reporter plasmids and β-galactosidase were lysed in the Passive Lysis Buffer (Promega). Both

the firefly (Fluc) and *Renilla* luciferase (Rluc) activities were determined using the Dual-Luciferase Reporter Assay kit as previously described (19). The firefly luciferase activity from monocistronic constructs was determined using the Mono-Luciferase Reporter Assay kit. The activity of  $\beta$ -galactosidase was measured as previously described (23) and was used to normalize the differences of transfection efficiency.

**RNA extraction and RNase Protection Assay.** Approximately  $2 \times 10^5$  H1299 cells in 10-cm plates were transfected with 4  $\mu$ g/plate DNA using Lipofectamine Plus reagent. At 48 hrs post-transfection, cells were harvested and the total RNAs were extracted using an RNeasy Mini Kit. Potential contaminating plasmid DNA in the total RNAs was digested with RNase-free DNase.

Ribonuclease Protection Assay (RPA) was performed using the RPA III kit according to the manufacturer's instruction. To prepare RNA probe, the region between -277 to -1 of human p27 5'-UTR sequence and +1 to +133 in the open reading frame of the firefly luciferase gene was amplified using PCR and cloned into pGEM-T Easy vector. The resulting plasmid was linearized with *Spe* I and labeled *in vitro* using T7 RNA polymerase in the presence of 0.5 mM each of ATP, UTP, GTP, and 0.01 mM CTP supplemented with 3.12  $\mu$ M  $\alpha$ -[ $^{32}$ P]CTP. The  $\alpha$ -[ $^{32}$ P]-labeled probe was digested with DNase and purified using a Sephadex G-25 Quick Spin Column. About  $5 \times 10^5$  c.p.m. of probe was hybridized to 15  $\mu$ g total RNA at 45°C overnight followed by digestion with RNase T1/A for 35 min at 37°C. The reaction was then stopped and the protected RNA samples were separated by electrophoresis on a 6% acrylamide/8 M urea gel. The gel was then dried for autoradiography.

## RESULTS

**The 5'-UTR sequence of p27 enhances expression of the second cistron in the dicistronic DNA test.** Dicistronic DNA test has been used in the previous studies to show the existence of IRES activities in the 5'-UTR sequence of both human and murine p27 although it was thought that the length of the IRES element may vary between the two species (8,9). To further characterize the 5'-UTR sequence, we engineered three dicistronic constructs containing various lengths of the 5'-UTR sequence of human p27 (Fig. 1B) flanked by *Renilla* and firefly luciferase genes (Fig. 1A). These constructs were then transfected into H1299 cells for analysis of potential IRES activity using luciferase assay. As shown in Fig. 1C, the expression of the second cistron firefly luciferase was greatly enhanced by the presence of the p27 5'-UTRs although the enhancement decreased gradually with the truncation of sequence from the 5'-end. The expression of *Renilla* luciferase was not changed by the insertion of the 5'-UTR sequence of p27. The increased ratio of firefly to *Renilla* luciferase activities by the 5'-UTR sequence of p27 compared to vector control suggests that there may be an IRES element in these sequences (Fig. 1D). These observations are consistent with the previous findings and with the claim of the existence of IRES activity in the 5'-UTR sequence of p27 (8-10).

We next engineered a hairpin loop in front of the *Renilla* luciferase gene of the dicistronic constructs to rule out the possibility that the enhanced expression of the second cistron was due to read-through or jumping from the first cistron (Fig. 2A). This hairpin structure has a free energy of -40 kcal/mol and it is expected to significantly inhibit the cap-dependent translation initiation of the first cistron (20). As shown in Fig. 2B and 2C, the enhanced expression of the second cistron firefly luciferase by the p27 5'-UTR was not affected by the hairpin structure. The expression of the first cistron *Renilla* luciferase, however, was significantly reduced compared with the constructs that do not have the hairpin (data not shown)

although the relative level of *Renilla* luciferase among the constructs with the hairpin did not change much (Fig. 2B). Thus, likely the ribosome read-through or jumping is not responsible for the enhanced expression of the second cistron firefly luciferase in the dicistronic DNA assay.

**The 5'-UTR of p27 does not display an IRES activity in the dicistronic mRNA assay.**

The above studies suggest that the 5'-UTR of p27 may contain an IRES element. However, the enhanced expression of the second cistron could also be due to the existence of promoters in the 5'-UTR or due to alternative splicing of the dicistronic transcripts as previously suggested (15-17). To further determine if the 5'-UTR of p27 contains IRES activities, we performed a more stringent dicistronic mRNA test which allows a direct analysis whether the 5'-UTR sequence of p27 in the intergenic region can increase the translation of the second cistron without the transcriptional and splicing interference. For this purpose, we engineered additional constructs for making *in vitro* dicistronic transcripts containing 5'-caps and 3'-polyA tails (Fig. 3A). In addition, a well known viral HRV IRES sequence was also cloned into the intergenic region to serve as a positive control. These *in vitro* transcripts were then transfected into H1299 cells which were then harvested for luciferase assay. As shown in Fig. 3B and 3C, the expression of the second cistron firefly luciferase was enhanced only by the HRV IRES as expected. None of the 5'-UTR sequences of p27 enhanced the expression of the second cistron firefly luciferase. Thus, it is likely that the 5'-UTR of p27 does not contain a functional IRES element as previously thought.

**The 5'-UTR sequence of p27 contains functional promoter elements.** The above results of dicistronic mRNA assay prompted us to investigate whether the 5'-UTR sequence of p27 contains a promoter which could be responsible for the observed stimulation of expression of the second cistron in the dicistronic DNA assays (Fig. 1 and Fig. 2). For this purpose, we first

engineered promoterless dicistronic constructs as described previously (20) by simply removing the unique SV40 promoter together with the intron sequence from the pRF-based dicistronic constructs (Fig. 4A). These promoterless dicistronic constructs will not generate dicistronic mRNAs after transfection into cells due to the lack of the SV40 promoter. Thus, the activity of any potential IRES element in the constructs should not be detected. Any expression of the second cistron firefly luciferase from these promoterless dicistronic DNAs will be due to the existence of a promoter in the 5'-UTR sequence of p27. As shown in Fig. 4B, the enhanced expression profile of the second cistron firefly luciferase in the promoterless vector is similar to that generated in the presence of the SV40 promoter (compare with Fig. 1 and Fig. 2), suggesting that the 5'-UTR sequence of p27 likely contains promoters.

To confirm the above conclusion, we further tested if the transcripts with only the second cistron firefly luciferase exist in the cells transfected with dicistronic DNA constructs. RNase protection assay was performed on total RNAs isolated from H1299 cells following transfection with the traditional and promoterless dicistronic constructs using a RNA probe covering the region between -277 to -1 of human p27 5'-UTR sequence and the region between +1 to +133 of firefly luciferase open reading frame (Fig. 5A). As shown in Fig. 5B, a short fragment with an estimated size of 124 bases was produced from the traditional dicistronic vector (lane 4) which represents the protected luciferase gene from the dicistronic RNA transcripts. This product is barely detectable with the RNA isolated from cells transfected with the promoterless dicistronic vector (lane 8), suggesting that the functional SV40 promoter has been successfully removed. A fragment with an estimated size of 415 bases was produced from the traditional dicistronic constructs containing the 575 and 461 bases of p27 5'-UTRs (lanes 5 and 6). This product is likely generated from the dicistronic RNA produced using the vector SV40 promoter. However,

a similar product, albeit with less intensity, was also observed with the promoterless dicistronic constructs (lane 9), suggesting that a transcription start site may exist upstream of the probe used (see the marker X in the diagram, Fig. 5A). Four fragments of 272, 258, 184, and 166 bases were also produced with both the traditional and the promoterless dicistronic constructs containing 575 and 461 bases of p27 5'-UTRs (lanes 5, 6, 9, and 10), suggesting that four transcription start sites likely exist at around -139, -125, -51, and -33. Interestingly, for the traditional dicistronic construct containing only 150 bases of the p27 5'-UTR (lane 7), only the product of 184 bases was observed, suggesting that except the transcription start site -51, all others are silent in this construct. The product of 290 bases (lane 7) was likely generated from the protection of the dicistronic RNA transcripts containing 150 bases of p27 5'-UTR. On the other hand, no significant protection of 290 and 184 bases was observed with the promoterless dicistronic constructs containing the 150 bases of p27 5'-UTR (lane 11), suggesting that the fragment of 290 bases is likely from the dicistronic RNA transcripts and the transcription start site at -51 requires the vector SV40 promoter and/or the upstream p27 promoters which may function as regulatory promoters (24). The later is supported by observation that the promoter activity for transcription initiation at -51 decreased with shortening of the 5'-UTR sequence of p27 from -575 to -150 (compare the intensity of the 184 and 166 base fragments in lane 11 with that in lanes 9 and 10).

To further confirm the promoter activities in the 5'-UTR sequence of p27, we cloned these sequences into the standard promoter-testing monocistronic vector pGL3 which does not have any promoter element. These constructs containing the different sizes of p27 5'-UTR were then transfected into H1299 cells for determination of luciferase activity. As shown in Fig. 6A, similar firefly luciferase activities were observed with these constructs compared to that generated from the traditional and promoterless dicistronic constructs (compare with Figs. 1, 2



and 4). RNase protection analysis of the RNAs isolated from cells transfected with these monocistronic constructs also showed the production of various transcription products similar to that shown with the dicistronic constructs (compare Fig. 6A with Fig. 5).

Based on the above observations with promoterless dicistronic and the standard promoter-testing monocistronic constructs, we conclude that the 5'-UTR sequence of human p27 contains promoter activities that can initiate transcription at several sites. Transcription at these sites in the dicistronic DNA assay generates monocistronic transcripts containing only the firefly luciferase gene with short 5'-UTRs that can be translated easily using the scanning mechanism (12,13). These shorter monocistronic RNA transcripts are likely responsible for the enhanced expression of firefly luciferase in the dicistronic DNA test and they may represent the previously reported IRES activity.

## **DISCUSSION**

In this work, we studied the 5'-UTR sequence of human p27 mRNA and further tested the putative IRES activity using promoterless dicistronic and monocistronic vectors and direct assay of dicistronic mRNAs. Our results indicated that the 5'-UTR sequence of p27 contains promoters which can drive production of abundant monocistronic transcripts with shorter 5'-UTRs. These findings raise questions on the IRES activity of the 5'-UTR sequence of p27 reported previously using the traditional dicistronic DNA assay.

Use of IRES for translating the untranslatable cellular mRNAs with long 5'-UTRs is an irresistible concept (25). The use of such a mechanism for translating polycistronic viral RNAs has been convincingly demonstrated previously (26). However, whether cellular mRNAs also use this mechanism is currently under debate (15,27). The central problem associated with the majority of studies claiming cellular IRES is the use of traditional dicistronic DNA test which

does not directly test the translation. Potential promoter activities and alternative splicing could both contribute to the observed "IRES" activity. These potential problems have recently been demonstrated on several previously believed-to-be cellular IRES elements. These IRES activities were shown to be due to cryptic promoters present in the 5'-UTRs (19,22,28,29) or due to differential splicing (30) which would create smaller monocistronic transcripts from the dicistronic DNAs for potential cap-dependent translation initiation of the second cistron. In one of these studies, the prevailing cellular IRES sequence of eIF4G (31,32) was shown to be an C/E BP beta transcription factor binding site (19). In another study, the strong IRES activity in the 5'-UTR of XIAP mRNAs was found to be mostly due to alternative splicing (30). Furthermore, a promoter activity was also found in the 5'-UTR sequence of hepatitis C virus which has previously been shown to have IRES activities (33). While these studies do not necessarily exclude in general the existence of IRES in cellular mRNAs *per se*, they necessitate more stringent studies such as tests using promoterless dicistronic DNA transfection, dicistronic RNA transfection, and RT-PCR for safeguarding any claims of cellular IRES.

Previously, the 5'-UTR sequences of both human and murine p27 were reported to contain IRES activities using traditional dicistronic DNA assays (8,9). While the IRES in the mouse p27 was thought to be located in the 217 bases proximal to the AUG start codon, the full IRES activity in human p27 was thought to be located within the 356 bases upstream of the AUG start codon. However, in the present study, we showed that these previously reported IRES activities are likely due to the promoters present in the 5'-UTR sequence of p27 (see Discussion below). In the case of human p27, the deletion analysis showed that the 184 bases upstream of the AUG start codon contains about half of the putative IRES activity with a 15 fold stimulation of Fluc/Rluc ratio by the 5'-UTR of p27 compared with the vector (9). This extent of stimulation

is similar to the promoter activity we found using the dicistronic construct containing the 150 bases of human p27 5'-UTR (see Discussion below). Furthermore, no stimulation of the second cistron was observed with the 5'-UTR of p27 in the dicistronic RNA assay (Fig. 3). In fact, the translation of the second cistron was inhibited in the dicistronic RNA assay by inserting the 5'-UTR sequence of p27 into the intergenic region. These observations further suggest that the 5'-UTR of human p27 may not contain any significant IRES element as previously thought.

Using both the promoterless dicistronic vector and the standard promoter-testing monocistronic vector pGL3, we demonstrated that the 5'-UTR of p27 contains promoters to stimulate the transcription of its following sequences. Because deletion of the first 114 bases reduced about half of the activity to stimulate the expression of firefly luciferase gene, it is likely that the 575 bases of the 5'-UTR sequence of p27 contain a promoter at the 5'-end which is absent in the construct harboring the 3' 461 bases of the 5'-UTR sequence. Indeed, we found by RNase protection assay that a transcription start site exists upstream of our probe (marked as X in Fig. 5A). Furthermore, Kullmann et al. (9) also found that the 575 bases of the human p27 5'-UTR contains a promoter which is absent in the 461 bases of p27 5'-UTR which is consistent with our findings. Further deletion at the 5'-end from -461 to -150 drastically reduced the activity in stimulating the luciferase expression. This observation suggests that the cryptic promoters likely exist in the region between -461 and -150. This conclusion is confirmed by RNase protection assay which showed that two transcription start sites at positions -125 and -139 are present in both the 575 and 461 constructs but they are missing in the -150 construct (Fig. 5). It is, however, noteworthy that the remaining 150 bases in the 5'-UTR of human p27 still stimulate the expression of the luciferase reporter albeit at a much reduced level compared with the longer sequences, suggesting that another promoter likely exists in the 150 bases upstream of

the AUG start codon. RNase protection assay confirmed that there is indeed a transcription start site at about -51 in all three constructs. However, this promoter activity appears to be very weak in the absence of upstream promoters such as the vector SV40 promoter or promoters present in the upstream sequence of p27 5'-UTR. It, thus, remains to be determined whether these putative promoters in chromosomes have any physiological roles in regulating p27 expression. We are currently testing this possibility and working toward this direction.

## References

1. Hengst, L. and Reed, S.I. (1998) *Curr Top Microbiol Immunol*, **227**, 25-41.
2. Sherr, C.J. and Roberts, J.M. (1999) *Genes Dev*, **13**, 1501-1512.
3. Millard, S.S., Yan, J.S., Nguyen, H., Pagano, M., Kiyokawa, H. and Koff, A. (1997) *J Biol Chem*, **272**, 7093-7098.
4. Agrawal, D., Hauser, P., McPherson, F., Dong, F., Garcia, A. and Pledger, W.J. (1996) *Mol Cell Biol*, **16**, 4327-4336.
5. Hengst, L. and Reed, S.I. (1996) *Science*, **271**, 1861-1864.
6. Keene, J.D. (1999) *Proc Natl Acad Sci U S A*, **96**, 5-7.
7. Millard, S.S., Vidal, A., Markus, M. and Koff, A. (2000) *Mol Cell Biol*, **20**, 5947-5959.
8. Miskimins, W.K., Wang, G., Hawkinson, M. and Miskimins, R. (2001) *Mol Cell Biol*, **21**, 4960-4967.
9. Kullmann, M., Gopfert, U., Siewe, B. and Hengst, L. (2002) *Genes Dev*, **16**, 3087-3099.
10. Gopfert, U., Kullmann, M. and Hengst, L. (2003) *Hum Mol Genet*, **12**, 1767-1779.
11. Sonenberg, N. (1994) *Biochimie*, **76**, 839-846.
12. Kozak, M. (1991) *J Biol Chem*, **266**, 19867-19870.
13. Kozak, M. and Shatkin, A.J. (1978) *Cell*, **13**, 201-212.
14. Merrick, W.C. (2004) *Gene*, **332**, 1-11.
15. Kozak, M. (2001) *Mol Cell Biol*, **21**, 1899-1907.
16. Kozak, M. (2003) *Gene*, **318**, 1-23.
17. Kozak, M. (2002) *Gene*, **299**, 1-34.
18. Stoneley, M., Subkhankulova, T., Le Quesne, J.P., Coldwell, M.J., Jopling, C.L., Belsham, G.J. and Willis, A.E. (2000) *Nucleic Acids Res*, **28**, 687-694.

19. Han, B. and Zhang, J.T. (2002) *Mol Cell Biol*, **22**, 7372-7384.
20. Han, B., Dong, Z., Liu, Y., Chen, Q., Hashimoto, K. and Zhang, J.T. (2003) *Oncogene*, **22**, 5325-5337.
21. Zhang, J.T. and Ling, V. (1991) *J Biol Chem*, **266**, 18224-18232.
22. Han, B., Dong, Z. and Zhang, J.T. (2003) *J Biol Chem*, **278**, 46983-46993.
23. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning*, New York.
24. Carey, M. (1998) *Cell*, **92**, 5-8.
25. Vagner, S., Galy, B. and Pyronnet, S. (2001) *EMBO Rep*, **2**, 893-898.
26. Pe'ery, T. and Mathews, M.B. (2000) In Soneberg, N., Hershey, J. W. and Mathews, M. B. (eds.), *Translational control of gene expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 371-424.
27. Schneider, R. and Kozak, M. (2001) *Mol. Cell. Biol.*, **21**, 8238-8246.
28. Hecht, K., Bailey, J.E. and Minas, W. (2002) *FEMS Yeast Res*, **2**, 215-224.
29. Verge, V., Vonlanthen, M., Masson, J.M., Trachsel, H. and Altmann, M. (2004) *Rna*, **10**, 277-286.
30. Van Eden, M.E., Byrd, M.P., Sherrill, K.W. and Lloyd, R.E. (2004) *Rna*, **10**, 720-730.
31. Gan, W., Celle, M.L. and Rhoads, R.E. (1998) *J Biol Chem*, **273**, 5006-5012.
32. Gan, W. and Rhoads, R.E. (1996) *J Biol Chem*, **271**, 623-626.
33. Dumas, E., Staedel, C., Colombat, M., Reigadas, S., Chabas, S., Astier-Gin, T., Cahour, A., Litvak, S. and Ventura, M. (2003) *Nucleic Acids Res*, **31**, 1275-1281.

## Figure Legends

**Figure 1. Dicistronic DNA test of the 5'-UTR sequence of human p27.** (A) Schematic diagram of dicistronic DNA constructs without (pRF) or with the insert of 5'-UTR of p27 (pR-P27-F) in the intergenic region. The locations of several relevant restriction enzyme sites are shown by arrows. (B) Schematic diagram of deletion of the 5'-UTR sequence of p27. The deletions were from the 5'-end of the sequence with the positions of deletion indicated. (C) Relative luciferase activity generated by the dicistronic DNA constructs. H1299 cells were transfected with pRF and pR-P27-F constructs. Twenty-four hours following transfection, cells were harvested and the *Renilla* and firefly luciferases activities were measured. Firefly (filled bars) and *Renilla* (open bars) luciferase activities were normalized to the co-transfected  $\beta$ -galactosidase activity and then to that generated by the vector (RF). (D) Relative F/R luciferase activity ratios. The relative ratios of firefly and *Renilla* luciferase activities were calculated and normalized to that of the vector-transfected cells (RF). The data were from four independent experiments. R.=*Renilla*; F.=firefly.

**Figure 2. Effect of hairpin structure on firefly luciferase expression by dicistronic DNA test.** (A) Schematic diagram of dicistronic DNA constructs containing a hairpin without insert (phRF) or with the 5'-UTR of p27 (phR-P27-F) in the intergenic region. The hairpin is shown between boxes representing the chimeric intron and the *Renilla* luciferase gene. (B) Relative luciferase activity generated by the dicistronic DNA constructs containing a hairpin. H1299 cells were transfected with phRF and phR-P27-F constructs. Twenty-four hours following transfection, cells were harvested and the *Renilla* and firefly luciferases activities were measured. Firefly (filled bars) and *Renilla* (open bars) luciferase activities were normalized to the co-transfected  $\beta$ -galactosidase activity and then to that generated by the vector (hRF). (C) Relative

F/R luciferase activity ratios. The relative ratios of firefly and *Renilla* luciferase activities were calculated and normalized to that of the vector-transfected cells (hRF). The data were from four independent experiments. R.=*Renilla*; F.=firefly.

**Figure 3. Dicistronic mRNA test of the 5'-UTR of human p27 in H1299 cells.** (A) Schematic diagram of the dicistronic mRNA used for translation in H1299 cells. *In vitro* transcripts with 5'-cap (m<sup>7</sup>GpppG) and 3'-polyA tail (A<sub>30</sub>) were synthesized using T7 RNA polymerase from linearized vector alone (RF<sub>A30</sub>), constructs containing the IRES of HRV (R-HRV-F<sub>A30</sub>) or the 5'-UTRs of p27 (R-P27-F<sub>A30</sub>). (B) Relative luciferase activity from dicistronic mRNAs in H1299 cells. H1299 cells were transfected with the dicistronic mRNAs, and 8 hours following transfection, *Renilla* and firefly luciferase activities were measured and normalized to that of vector control (RF). (C) Relative F/R luciferase activity ratios. The ratios of relative firefly and *Renilla* luciferase activities were calculated and normalized to that of the control vector-transfected cells (RF).

**Figure 4. Promoterless dicistronic DNA test of the 5'-UTR sequence of human p27.** (A) Schematic diagram of promoterless dicistronic DNA constructs without insert (pRF(-P)) or with the 5'-UTR of p27 (pR-P27-F(-P)) in the intergenic region. The locations of several relevant restriction enzyme sites are shown by arrows. (B) Relative luciferase activity generated by the promoterless dicistronic DNA constructs. H1299 cells were transfected with pRF(-P) and pR-P27-F(-P) constructs. Twenty-four hours following transfection, cells were harvested and the *Renilla* and firefly luciferases activities were measured. Firefly and *Renilla* luciferase activities were normalized to the co-transfected  $\beta$ -galactosidase activity and then to that generated by the vector (RF). The relative ratios were calculated and normalized to that of the vector-transfected cells (RF(-P)). The data were from four independent experiments. R.=*Renilla*; F.=firefly.

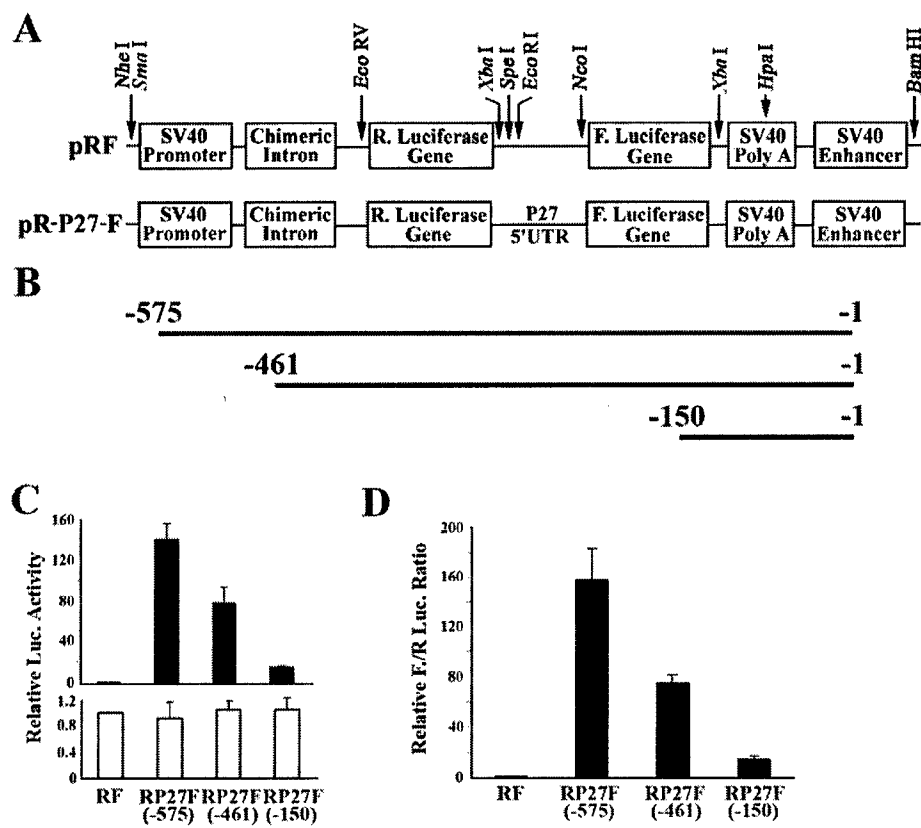


**Figure 5. RNase protection assay of p27 5'-UTR in dicistronic constructs.**

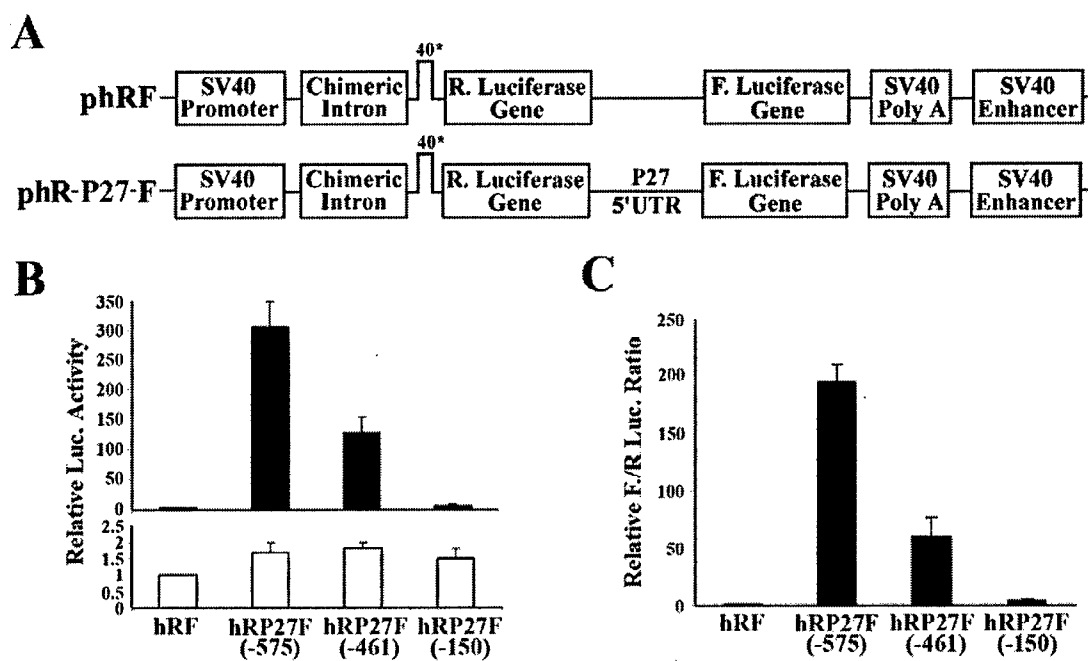
(A) Schematic diagram illustrating the location of transcription start sites and the probe used for RNase protection assay. X indicates an upstream transcription start site. (B) RNase protection assay. Total RNAs were isolated from H1299 cells following transfection with dicistronic vector pRF (lane 4), pR-P27-F (lanes 5-7), promoterless dicistronic vector pRF(-P) (lane 8), and promoterless pR-P27-F (lanes 9-11) and then used for RNase protection assay as described in Materials and Methods. Lanes 1-3 are control reactions without (lanes 1 and 2) or with (lane 3) RNase in the absence (lane 1) or presence (lanes 2-3) of yeast RNA.

**Figure 6. Promoter activity of p27 5'-UTR in monocistronic constructs and RNase protection assay.** (A) Schematic diagram illustrating the location of transcription start sites and the probe used for RNase protection assay. X indicates an upstream transcription start site. (B) Luciferase assay. H1299 cells were transfected with vector pGL3 and monocistronic constructs that harboring the 5'-UTR sequences of p27. Twenty-four hours following transfection, cells were harvested and luciferase activities were measured. The luciferase activities were then normalized to the co-transfected  $\beta$ -galactosidase activity and then to that generated by the vector (pGL3). (C) RNase protection assay. Total RNAs were isolated following transfection with the vector pGL3 (lane 1) and constructs containing the 5'-UTR sequences of p27 (lanes 2-4) and then used for RNase protection assay as described in Materials and Methods.

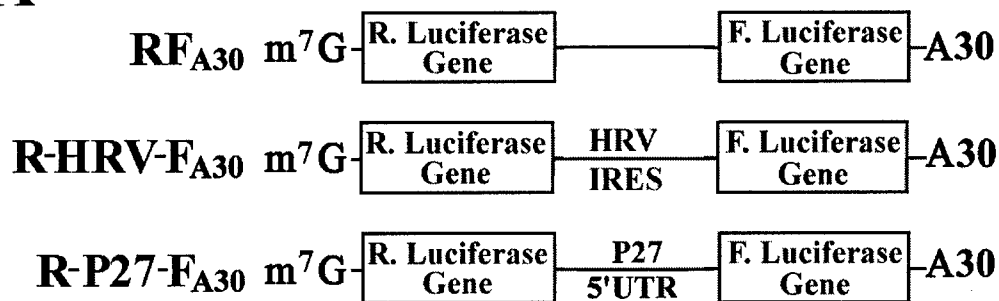
Liu et al., Figure 1



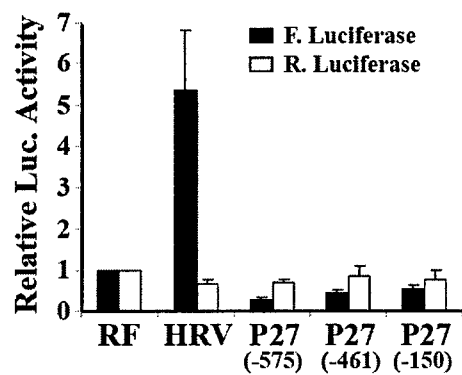
Liu et al., Figure 2



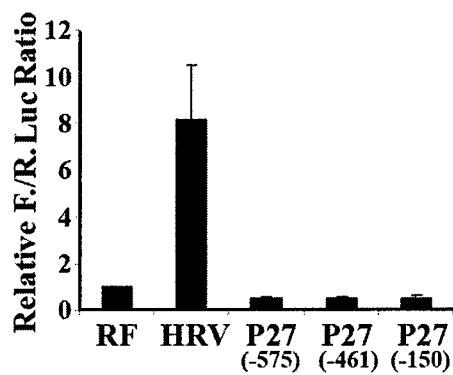
**A**

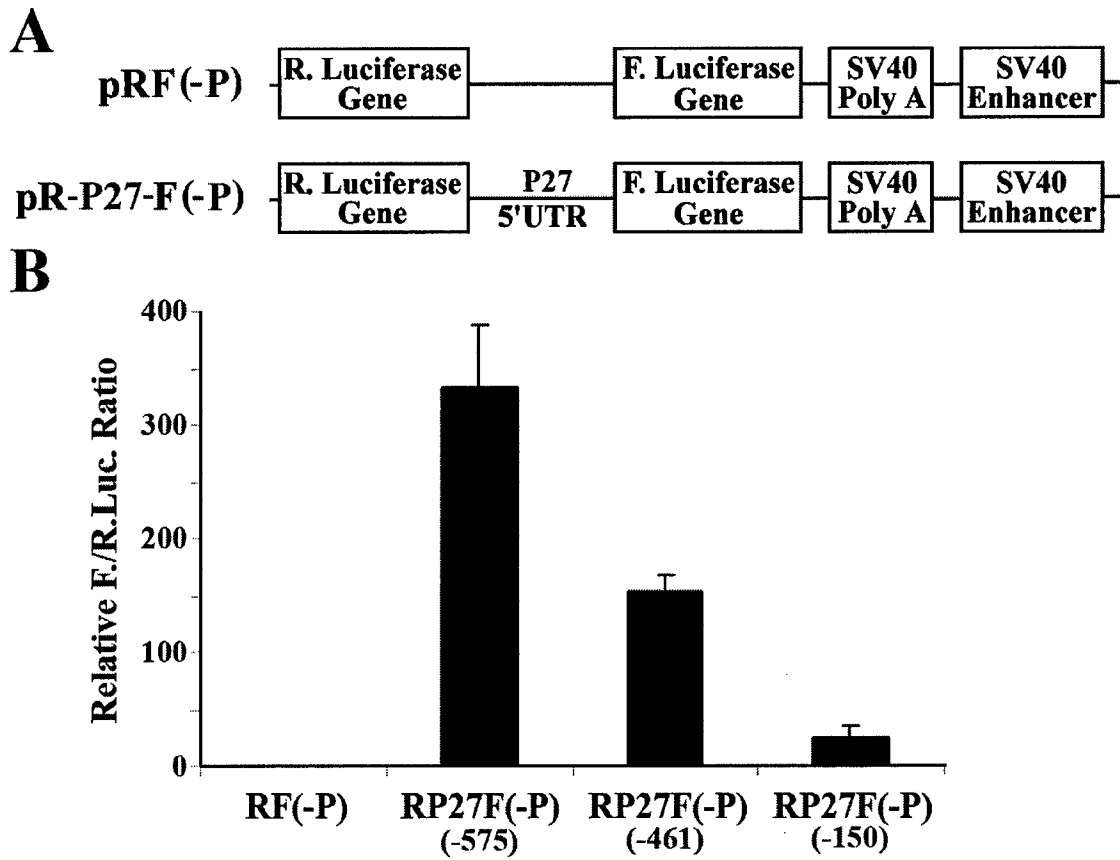


**B**

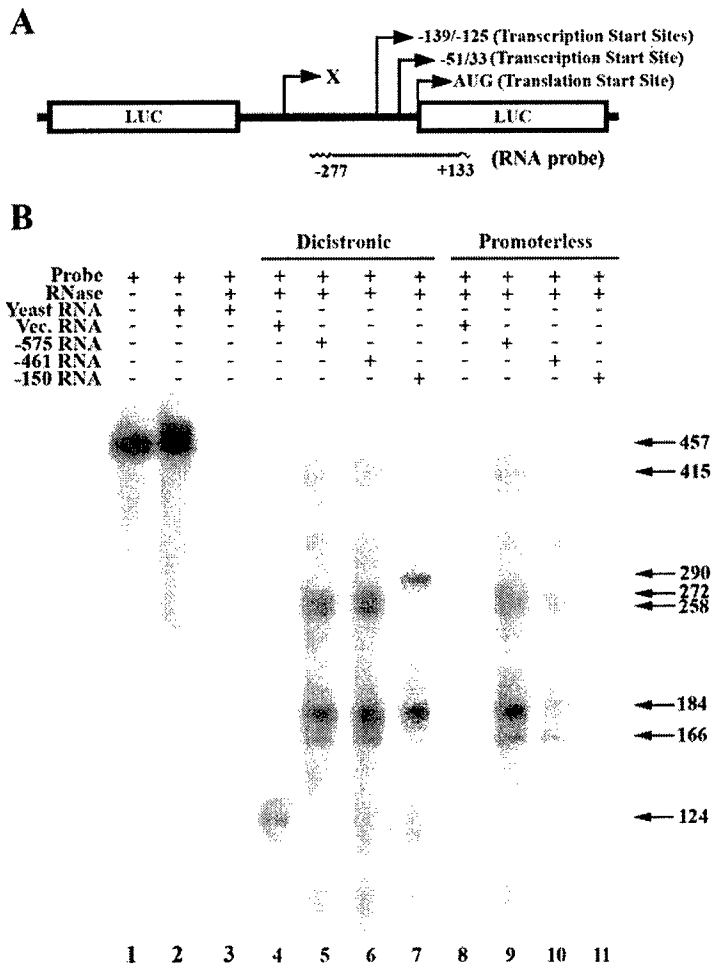


**C**





Liu et al., Figure 5



Liu et al., Figure 6

